Cloning and molecular characterization of a polygalacturonase gene of the white-rot fungus *Chondrostereum purpureum*

by

Yijian Tang
B.M., Hunan Medical University, 1990

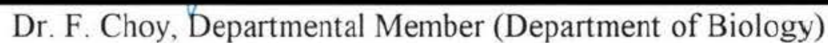
A Thesis Submitted in partial Fulfillment of the Requirement for the Degree of

MASTER OF SCIENCE

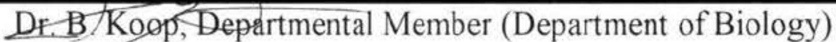
in the Department of Biology

We accept this thesis as conforming to the required standard

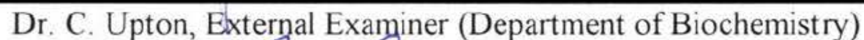
Dr. W. E. Hintz, Supervisor (Department of Biology)



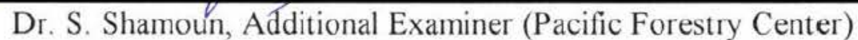
Dr. F. Choy, Departmental Member (Department of Biology)



Dr. B. Koop, Departmental Member (Department of Biology)



Dr. C. Upton, External Examiner (Department of Biochemistry)



Dr. S. Shamoun, Additional Examiner (Pacific Forestry Center)

© Yijian Tang, 1997

University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisor: Dr. William E. Hintz

ABSTRACT

The basidiomycete *Chondrostereum purpureum* is a white-rot fungus which is pathogenic to many hardwood species. We are interested in determining what factors make this fungus an effective pathogen.

Certain steps are crucial to the establishment of fungal pathogenicity. For the pathogen *C. purpureum*, penetration of the host and colonization of the host tissue seem to be especially important to the disease establishment. The plant cell wall is central to plant pathogens as both a barrier to infection and a source of metabolizable substrates. Cell wall degrading enzymes (CWDEs) have been implicated in the pathogenicity of certain phytopathogenic fungi. Endopolygalacturonase (endoPG) is a key CWDE in that it breaks down peptic polymers, a major wall component of plant cell. The endoPG produced by *C. purpureum* is able to cause the silver-leaf symptoms of infected fruit trees and endoPG purified from culture filtrates of *C. purpureum* has been shown to cause necrosis of leaf disc. A positive correlation between endoPG secretion level and the virulence of the fungus among different strains of *C. purpureum* has also been reported.

Pathogenicity is usually a multigenic trait, assessing the role of one potential pathogenicity factor requires a refined approach. This work takes the first step in a refined assessment of the role endoPG plays in the pathogenicity of *C. purpureum* by the

cloning and characterizing the endoPG gene. This will enable the construction of strains of *C. purpureum* which contain either (*i*) a disrupted endoPG gene such that no functional enzyme is produced or (*ii*) multiple copies of endoPG gene such that the enzyme will be overexpressed. Pathogenicity tests with these recombinant strains possessing essentially identical genetic backgrounds except for the number of endoPG genes and differing only in their secretion of this enzyme would clarify the role endoPG plays in the pathogenicity of *C. purpureum*.

Using RT-PCR with total RNA isolated from *C. purpureum* and degenerate primers designed from conserved regions from the multiple endoPG peptide alignment of bacteria, fungi and plants, a DNA fragment bearing significant similarity with other fungal endoPGs was amplified. This was used as a probe to screen the genomic library of *C. purpureum* and a putative endoPG gene was isolated from the library. The genomic copy of the gene was entirely sequenced. Furthermore, three overlapping cDNA fragments of the gene were amplified by RT-PCR using primers designed from the structural gene and the cDNAs were sequenced. These cDNA fragments span an open reading frame of 1083 bps. Sequence comparison of the structural gene and the cDNA showed that the gene consists of a promoter region sequenced up to -1592 bp, a coding region of 1596 bps which is interrupted by nine introns, and a terminator region of 108 bps. The deduced peptide sequence, 361 a.a. in length, has a polygalacturonase active site, confirming that the gene is a polygalacturonase encoding gene. Sequence comparison of the deduced peptide with the other PGs revealed the highest percentage of sequence similarity with fungal endoPGs whereas the percentage of its sequence similarity with fungal exoPGs dropped significantly. This suggests that the cloned PG

gene with *C. purpureum* genomic DNA digests showed only one hybridization signal. RT-PCR with primers designed from the genomic sequence of the gene amplified only one copy of cDNA molecules. The above indicate that a single copy of endoPG encoding gene exists in *C. purpureum* genome.

Having obtained the endoPG gene and the peptide sequence deduced from the cDNA, the transformation of the fungus for the endoPG gene disruption and gene overexpression becomes feasible.

Examiners:

Dr. W.E. Hintz, Supervisor (Department of Biology)

Dr. F. Choy, Departmental Member (Department of Biology)

Dr. B. Koop, Departmental Member (Department of Biology)

Dr. C. Upton, External Examiner (Department of Biochemistry)

Dr. S. Shamoun, Additional Examiner (Pacific Forestry Center)

Table of Contents

Abstract	ii
Table of Contents	v
List of Figures	vii
List of Tables	viii
List of Abbreviations	ix
Acknowledgement	x
Dedication	xi
INTRODUCTION	
1.0 Biological control in forestry vegetation management	1
2.0 *Chondrostereum purpureum* as a biological control agent	4
3.0 The role of endoPG in the pathogenicity of plant disease	6
3.1 EndoPG as a key cell wall degrading enzyme	7
3.2 EndoPG in fungal phytopathogenicity	10
3.3 Alternate role of endoPG in host-pathogen interaction	11
3.4 EndoPG in the pathogenicity of *Chondrostereum purpureum*	12
4.0 Summary	14
MATERIALS AND METHODS	
1.0 Fungal Culture	15
2.0 Nucleic acid extraction	15
3.0 Amplification of an endoPG gene fragment as a probe by RT-PCR	16
3.1 Design of the degenerate PCR primers from multiple sequence alignment	16
3.2 RT-PCR strategies	17
3.3 Cloning and sequencing	19
4.0 Southern hybridization	20
5.0 Genomic library construction and screening	21
6.0 Subcloning	21
7.0 Sequencing strategies for the genomic copy of the endoPG gene	23
8.0 cDNA amplification by RT-PCR	28
8.1 Primer design	28
8.2 RT-PCR strategies	30

8.3 Sequencing cDNA fragments 32

9.0 Phylogenetic analysis of polygalacturonases 33

RESULTS

1.0 Cloning of the endoPG gene from *C. purpureum* genomic DNA library 37

1.1 PCR cloning of a fragment of endoPG gene 37

1.2 Genomic library construction and screening 42

1.3 Sequencing of the genomic copy of the gene 43

2.0 Amplification of the cDNA fragments of the endoPG gene by RT-PCR 45

2.1 Amplification of the endoPG cDNA fragment1 46

2.2 Amplification of the endoPG cDNA fragment2 48

2.3 Amplification of the endoPG cDNA fragment3 49

2.4 EndoPG cDNA sequence and its deduced peptide sequence 51

3.0 Relatedness of polygalacturonase among bacteria, fungi and plants 51

DISCUSSION

1.0 Characterization of the endoPG gene 54

1.1 Promoter and 5' untranslated region of the gene 54

1.2 The coding region of the gene 57

1.3 The 3' untranslated region of the gene 60

2.0 One gene and at least four isoforms of the enzyme 60

3.0. Putative endoPG primary structure analysis, comparison with other PG sequences and secondary structure predictions 62

4.0. Further considerations for the study of the role of endoPG in the pathogenicity of *C. purpureum* 69

BIBLIOGRAPHY 71

APPENDIX 1: Results of multiple peptide sequence alignment of thirty-four polygalacturonases using PILEUP 78

APPENDIX 2: Towards transformation of *Chondrostereum purpureum* 92

List of Figures

Figure 1: The possible structure of pectic polymers Rhamnogalacturonan and covalently linked neutral polysaccharides 9

Figure 2: Cleavage of an α-1,4-linked polygalacturonide chain by endoPG and other pectic enzymes 9

Figure 3: Design of degenerate PCR primers from multiple sequence alignment 16

Figure 4: RT-PCR amplification of an endoPG gene fragment 18

Figure 5: Gene cloning outline 22

Figure 6: PCR used to determine the insert orientation in the subclone pePGsal6 and to estimate how much of the gene was contained in the clone 24

Figure 7: Sequencing strategy 27

Figure 8: cDNA amplification by RT-PCR 29

Figure 9: Alignment of the peptide sequence of the putative endoPG gene fragment from *C. purpureum* with endoPGs from fungi, plants and bacteria 40

Figure 10: Gene copy analysis by Southern hybridization 41

Figure 11: Gel photos of RT-PCR amplified endoPG cDNA fragments 47

Figure 12: Nucleic acid sequence of the cDNA of the polygalacturonase gene from *Chondrostereum purpureum* and the deduced amino acid sequence of the encoded protein 52

Figure 13: Dendrogram of relatedness of polygalacturonases from bacteria, fungi and plants 53

Figure 14: Nucleic acid sequence of the endoPG gene from *Chondrostereum purpureum* 55

Figure 15: The best conserved amino acid domain in PG sequences 66

Figure 16: Secretion signal peptide prediction for the *C. purpureum* endoPG gene product 68

List of Tables

Table 1: Sequence comparison of the 9 introns of the *C. purpureum* endoPG gene 59

Table 2: Codon usage of endoPG gene of *C. purpureum* 61

Table 3: Comparison of the amino acid composition of the putative product of the endoPG gene with the published endoPG I of *C. purpureum* 63

List of Abbreviations

%	percent
°C	degree centigrade
μg	microgram
μl	microlitre
3'	three prime
5'	five prime
bp	base pair
C-terminal	carboxyl-terminal
C. purpureum	*Chondrostereum purpureum*
cDNA	copy deoxyribonucleic acid
cm	centimetre
dATP	deoxyadenosine triphosphate
DNA	deoxyribonucleic acid
Dnase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
d(T)	deoxythymidine
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
endoPG	endopolygalacturonase
kDa	kilodalton
ml	millilitre
mM	millimolar
N terminal	amino-terminal
ng	nanogram
OD	optical density
PCR	polymerase chain reaction
PG	polygalacturonase
polyA	polyadenylated
RNA	ribonucleic acid
RT-PCR	reverse transcription linked polymerase chain reaction
SDS	sodium dodecyl sulphate
TAE	tris acetate EDTA
Taq	*Thermus aquaticus*
TBE	tris borate EDTA
UV	ultraviolet

Acknowledgment

I would like to thank my supervisor Dr. Willam Hintz for giving me the opportunity and support, Dr. Francis Choy and Dr. Simon Shamoun for encouragement, Dr. Ben Koop for much help with the peptide alignment and for serving on my committee at the last minute instead of a former committee member Dr. Pearson who is on sabbatical leave in Africa now.

Dr. Petr Karlovsky designed and provided the degenerate primers. His advices were invaluable to me throughout the whole work. He also spent a lot of time helping with the phylogenetic analysis of the 34 polygalacturonases and with the drawing of the figures and tables of my thesis.

Life here became much more pleasant because of these friends: Caihua who is great both at work and off work, Kaiyu, Cindy, Manish, Mark and Bev.

Nothing would be possible without Xiaoyang, who showed me the way eight years ago.

Dedication

This thesis is dedicated to my father, my mother, my sisters and my brother, for all the love and support they have given me and for their believing in me.

Introduction

1.0 Biological control in forest vegetation management

Vegetation management in Canadian forests requires an efficient method for the control of unwanted rapidly growing weed species. Hardwood trees, such as *Alnus rubra* Bong (red alder), *Populus nigra* (black cottonwood), *Populus tremuloides* (Aspen poplar), *Betula* sp.(birch) and *Acer spp.* (bigleaf maple and Douglas maple) are considered to be weeds in certain forest applications. In forest renewal sites, they compete with newly planted or naturally regenerated trees, such as conifers, which are of high commercial value. The fast-growing hardwoods compete for resources and shade the high-value conifers. Hardwoods are also a problem in hydro-electric rights-of-way (R-O-W), because they encroach upon electric transmission lines and may cause fire hazard. Manual cutting of the competing vegetation (forest weeds) is costly and is often rendered ineffective by the regrowth of sprouts from the cut stumps. In a continued vegetation management program frequent return visits are needed. Mechanical or incendiary site preparation methods are successful on many sites but a common problem is the rapid re-establishment of weed species from rhizomes or windblown seeds (Wilson, 1968; Kellman, 1970; Archibold, 1979; Oswald, 1990). To date the most effective and economical way for the suppression of forest weeds is by manual cutting followed by application of chemical herbicides. The disadvantage of using chemical herbicides is that being synthetic products they are difficult to biodegrade and can seep through soil and

water. Generally, chemical herbicides act by blocking the metabolic pathways in the weed species but little is known of their effects on non-target organisms hence there is a perceived risk to humans. Potential side-effects have thus caused public resistance to their use. Chemical herbicide development is also faced with the need for more target specific chemicals at escalating development costs (Braunholto, 1981; Hill, 1982; Shaw, 1978). It seems possible that the opportunities for registering new chemicals for forestry use are rapidly diminishing and the likelihood of de-registration of existing products is increasing (Wall, R. E., 1990). Clearly there is a need for research on alternative vegetation management methods which are environmentally acceptable and economically feasible.

Biological control, which is defined as the deliberate use of natural enemies to suppress the growth or reduce the population of a weed species (Watson, 1989), provides an attractive alternative. It is generally accepted that for every species there is at least one direct competitor or pathogen which serves to strike a balance in the ecosystem. In theory it should be possible to find at least one organism in the natural environment that can be used for the biocontrol of any other organism. Present regulations for the registration of biological control agents allow the use of indigenous species within defined geographic zones only. This ensures that the test organism is itself subject to natural controls and will neither persist in the environment at greater than endemic levels nor spread appreciably from the treated site. Another advantage for the use of biological control is that field applied material is bio-degradable and thus will not pollute soil or water (Wall et al., 1992).

Many of the pathogens causing disease of woody plants are fungi hence it logically follows that fungi are being developed as biocontrol agents to suppress forest weeds in specific forest applications such as forest renewal sites. Mycoherbicides are defined as formulations containing indigenous phytopathogenic fungi that are applied as an inundative inoculum to target weeds. Certain fungi have already been used as practical tools for biocontrol in agricultural practices. Examples of useful phytopathogenic fungi include: *Phytophthora palmivora* Butler which is marketed as DEVine™ for the control of strangler vine in citrus groves in the Southern United States (Ridings, 1986), *Colletotrichum gloeosporiodes* (Penz.) Sacc. f. sp. *aeschynomene* marketed under the trade name Collego™ for the control of northern joint-vetch in rice and soybean fields (Templeton et al., 1984), and *Colletotrichum gloeosporioides f.sp.malvae* marketed as Biomal to control round-leafed mallow in wheat and lentil crops in Manitoba and Saskatchwan (Mortensen, 1988; Auld, 1990). The use of mycoherbicides for integrated vegetation management in forestry can be more attractive than in agricultural applications as the inoculation can be timed to take advantage of environmental conditions favorable for disease development without the constraints imposed by crop management (Templeton, 1986; Wall, et al., 1992).

The general scenario for the development of a mycoherbicide proceeds from discovery of the diseased weed species in the field. The causal fungus is cultured on artificial medium, identified and reinoculated on the weed host. Successful reproduction of the disease and reisolation of the fungus from the weed host establishes the fungus as a potential mycoherbicide. Further efficacy testing includes infectivity and pathogenicity

trials under various environmental conditions, small scale field tests, host range studies and evaluation of field survival. If these tests indicate that an effective product can be developed, further steps must then be taken for its registration as a mycoherbicide. Prior to marketing additional research includes the development of mass culture techniques, testing for toxic or allergenic effects on animals, non-target effects and determining the environmental fate of the released organism (Templeton et al., 1980).

2.0 *Chondrostereum purpureum* as a biological control agent

Since the target species of vegetation management in B.C. forest and R-O-W applications are mixed hard woods, it is desirable to find a fungal pathogen having broad host range for development as a mycoherbicide. The basidiomycete *Chondrostereum purpureum* (Pers. ex Fr.) Pouzar provides an attractive candidate. This fungus is a pathogen of dicotyledonous trees and shrubs throughout the temperate zones of the world (Peace, 1962). It is grouped in the white-rot fungi and being a stem pathogen it can debilitate target plants relatively quickly by disrupting the vascular tissues thereby effectively severing the physiological connection between foliage and roots. If the fungus invades the cambium of the lower stem or the root crown, it may also prevent regrowth of adventitious shoots. *Chondrostereum purpureum* usually invades woody angiosperms through pruning wounds, cut stumps or stem lesions and causes cambial necrosis, sapwood stain and decay, foliar discoloration and eventually host death (Rayner, 1977; Bishop, 1979; Wall, 1986). It has been shown that treatment of hardwood cut stumps with cultures of *C. purpureum* caused a reduction in survival of stump sprouts

(Scheepens and Hoogerbrugge 1988, Wall 1990). Inoculation of the fungus through axe wounds and frilling on yellow birch and red alder resulted in the suppression of wound healing and cambial recovery as well as elongate cankers. There was also an increase in the likelihood of mortality in the case of application of the fungus to the complete frills (Wall 1986, 1990). A precedent for using *C. purpureum* as mycoherbicide has been set in the Netherlands where Scheepens and coworkers (1980) obtained control of stump sprouting of American black cherry (*Prunus seritina* Ehl.) trees by inoculating cut stumps with *C. purpureum*.

Disease Cycle:

Wood rotting fungi enter trees as germinating basidiospores or mycelium through wounds, dead branches, branch stubs, tree stumps or damaged roots and spread from the point of entry to the heartwood and/or sapwood of the tree or tree sprouts. Wounds caused by fire and by pruning operations are the most common points of entry of these fungi. Following successful invasion, the fungus develops in the wood and spreads upward and/or downward throughout the tree eventually being manifested as disease symptoms. The process of discoloration and decay in the wood of living trees is quite complex. Following an injury to the tree, the injured cells and those surrounding them undergo chemical changes such as oxidation and often become discolored. A variety of microorganisms can land on the surface of a tree wound and begin to grow on the moist nutrient-rich surface. The wood-rotting Basidiomycetes attack only those tissues that have first been altered by the chemical processes, bacteria and other fungi including the

Ascomycetes and imperfect Fungi. They grow intrusively and utilize the cell wall components for food and energy, reducing the wood to a light-colored spongy mass (white rot) or cracking brown mass (brown-rot). Of the two categories *C. purpureum* is grouped with the white-rot fungi. The sporophores or conks of wood-rotting basidiomycetes appear near the point of entry of the fungus, in cankers or swollen knots along the stem of living trees, or along the length of the tree stem following its death. The sporophores produce basidiospores during the growing season and the spores are carried by wind, rain or animals to the nearby trees (Agrios, 1978).

Chondrostereum purpureum shows great promise for development as a mycoherbicide for the control of mixed hardwood weeds. Prior to its registration as a mycoherbicide in Canada, more in-depth research in the pathogenicity study, risk assessment and environmental fate of the fungus is required.

Purpose of this research:

The determination of the combined factors which make *Chondrostereum purpureum* an effective pathogen is of both academic and practical value. This current reseach takes the first step in determining the role of a potential pathogenicity factor, a cell wall degrading enzyme endopolygalacturonase, in the pathogenicity of *Chondrostereum purpureum*. The knowledge obtained from a pathogenicity study has an immediate application: to aid in strain optimization in the development of the fungus as a mycoherbicide.

3.0 The role of endopolygalacturonase in the pathogenicity of plant disease

The range of phytopathogenic organisms that attack plants is diverse and includes viruses, mycoplasma, bacteria, fungi, nematodes, protozoa, and parasites. Despite the vast array of potential phytopathogens, plant resistance is the rule and susceptibility is the exception. Of the 100,000 known species of fungi, only 8000 can cause disease on one or more plant species (Agrios, 1988). In the gene-for-gene hypothesis of the plant-pathogen interaction, the recognition of a pathogen triggers defense mechanism in the host; since the recognition event is a specific binding of an elicitor from the pathogen to a receptor in the host cell, the war between a plant and its pathogen is believed to be, for the pathogen to lose the elicitor molecule, and for the plant to come up with a new receptor that can recognize the pathogen again. The evolutionary process is thus a move-counter-move process, as is in fact any process of host-pathogen interaction. In view of the evolutionary process of the plant-pathogen interaction, the possible existence and the nature of fungal pathogenicity mechanisms are inseparable from the plant's ability to recognize a foreign organism and set up defense mechanisms. Some steps that may be crucial to the establishment of fungal pathogenicity are: attachment to the plant surface; germination on the plant surface and formation of infection structures; penetration of the host and colonization of the host tissue (Schafer, 1994). For the pathogen *C. purpureum*, penetration of the host and colonization of the host tissue may be especially important to the disease establishment because *C. purpureum* is known to require a preexisting wound surface on the trees for its entry.

3.1 Endopolygalacturonase as a key cell wall degrading enzyme

The plant cell wall is central to plant pathogens as both a barrier to infection and a source of metabolizable substrates. Most, if not all cellular pathogens possess an array of enzymes that can degrade plant cell wall polymers. The role of cell wall-degrading enzymes (CWDE) in disease establishment and symptom development has been the subject of extensive research as well as of much controversy (Bateman and Basham, 1976; Cooper, 1984; Hahn et al., 1989). The plant cell wall is composed of pectic polymers, hemicelluloses and cellulose with intimate yet not completely known association between different polymers. Pectic polymers, constituting about 35% of primary walls of dicots, are predominant in but not exclusive to the intercellular region or middle lamella. This zone is of particular importance in parasitism as it is exploited by many pathogens which grow between cells. The key polymer, rhamnogalacturonan, is comprised of chains of α-1,4-linked galacturonic acid residues interspersed with 1,2-linked rhamnose (Fig. 1). Rhamnogalacturonan is degraded by hydrolytic polygalacturonases (PG) and lyases (PL) (Fig. 2). Both may attack the chain internally at random (endo-) or terminally (exo-) but some enzymes combine the two modes of action and attack chains at random followed by the release of mono- or oligomers (Cooper et al., 1978).

The key role of this polymer in plant cell wall structure is revealed by the drastic effects caused by endopolygalacturinidases which not only solubilize galacturonides but result in a marked depletion of the covalently linked neutral sugars, arabinose, galactose and rhamnose (Basham and Bateman, 1975; Talmadge et al., 1973). Cooper (1983) has

summarized and reviewed evidence for the involvement of these enzymes which randomly degrade rhamnogalacturonan in pathogenesis by facultative parasites. When certain pathogens are grown on insoluble extracted host cell walls, the first cell wall degrading enzymes produced *in vitro* are endopolygalacturonases.

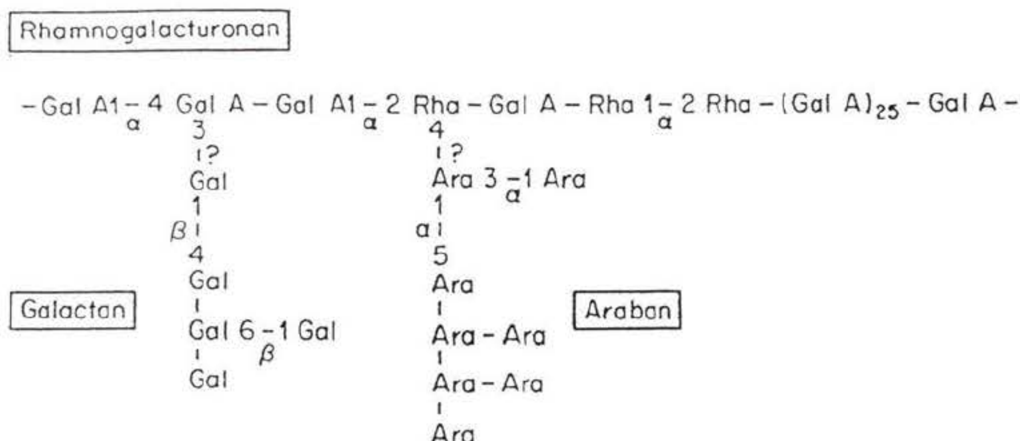


Figure 1: Possible structure of pectic polymers Rhamnogalacturonan and covalently linked neutral polysaccharides.

GalA, galacturonic acid; Rha, rhamnose; Gal, galactose; Ara, arabinose. (Albersheim, 1976; Darvill et al. 1980).

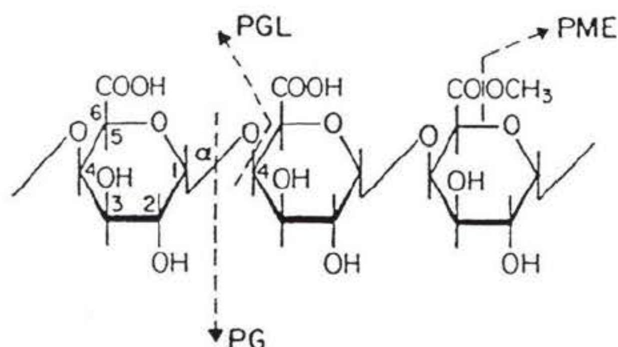


Figure 2: Cleavage of an α-1,4-linked polygalacturonide chain by endopolygalacturonase (PG) and endopolygalacturonide lyase (PGL), and demethoxylation by pectin methylesterase (PME)

Note the numbering of carbons on the galacturonosyl units, the α configuration of the glycosyl bond, and the methylated or free carboxyls.

Under these conditions cell wall degrading enzyme production is sequential in the order of pectic enzymes, hemicellulases, cellulases. This progression seems to reflect the physio-chemical susceptibilities of corresponding wall polymers. Endopolygalacturonidases thus seem to be key enzymes in the breaking down of the cell wall (Cooper, 1977;1983). Studies with purified endoPGs and PLs from many plant pathogenic bacteria and fungi have clearly demonstrated their unique ability to cause maceration (cell separation) and simultaneous damage to protoplasts while exo-acting PGs and other CWDE are ineffective (Bateman & Basham, 1976). In addition, numerous side-effects may result from degradation of host rhamnogalacturonan and give rise to diverse symptoms and damage since cell wall bound molecules with potential physiological or damaging effects may be exposed or released following wall breakdown (Cooper, 1978;1983).

3.2 Endopolygalacturonase in fungal phytopathogenicity

Pathogenicity studies of certain fungi have implicated endopolygalacturonase as a virulence factor (Mckeen, 1974; Mansfield and Richardson, 1981; Cleveland et al., 1991; Cervone et al., 1987). EndoPG enzymes have been visualized, using a PG-binding protein as a probe, in the tissue area where extensive pectin breakdown or cell separation occured in bean infected by *Colletotrichum lindemuthianum* (Benhamou et al., 1991) and in soybean treated with an endoPG by *Sclerotinia sclerotiorum* (Favaron et al., 1993). Convincing evidence of endoPG involvement in the pathogenesis has been obtained for *Aspergillus flavus*. Low-virulence strains lacking a major endoPG caused less severe disease symptoms in developing cotton bolls than one high-virulence strain that produced

a major non-catabolite repressed endoPG (Brown et al., 1992; Cleveland and Cotty, 1991). On the whole, genetic evidence for the involvement of endoPG in the fungal disease of plants is still limited and several of such studies relied on the analysis of chemically induced mutants. A more refined and preferred approach is the disruption of specific putative pathogenicity loci by transformation-mediated site-specific mutagenesis. Disruption of the polygalacturonase gene of *Cochliobolus carbonum*, however, demonstrated that endoPG-minus fungal strains were still effective pathogens (Scott-Craig, et al, 1990). Simple incapacitation of this single locus had little or no effect on overall pathogenicity but it should be remembered that pathogenicity is usually a multiallelic trait. It is also worth mentioning that in *Cochliobolus* case the host plant (maize) was a monocotyledon and contained much less pectin, 8-9% as compared to 35% in dicots. We may reasonably expect a more important role for endoPG in pathogenesis toward dicots such as hardwood trees. In general, however, there is insufficient evidence to generalize a correlation between a pathogen's success and the levels of cell wall degrading enzymes secreted by different pathogens.

3.3 Alternate roles of endoPG in host-pathogen interaction

Host-pathogen interactions can be very complex and the role of endoPGs must be considered in the context of host-pathogen co-evolution. Interactions include: (1) The secretion of endoPG by pathogens as one of their chemical weapons during invasion to plants. (2) Cell wall fragments released by endoPG can have regulatory functions on the host plants, including elicitation of defense mechanisms (as reviewed by Albersheim et.

al. 1982). (3) Plants can secrete endoPG-inhibiting protein to favor the accumulation of elicitor-active fragments - oligogalacturonides with a degree of polymerization between 10 and 14, which have been shown to induce phytoalexin accumulation and lignification (Cervone et al , 1987; 1989; Fielding 1981). It has been recently suggested that there has been an evolution of PG toward less efficient enzymatic forms during the conversion of saprophytes into facultative parasites and finally into obligate parasites. A highly efficient enzyme rapidly forms oligogalacturonide fragments which are capable of eliciting a resistance response whereas a less efficient enzyme may delay elicitation of a resistance response allowing colonization of the tissue (Cervone, 1986). Fungal phytopathogens that utilize endoPG to invade and colonize plant tissue would not be expected to lose the function of these enzymes and thus still remain phytopathogens.

3.4 Endopolygalacturonse in the pathogenicity of *Chondrostereum purpureum*

The involvement of endoPG in the pathogenesis of *C. purpureum* has been suggested by previous studies. *Chondrostereum purpureum* is similar to other wood decaying basidiomycetes of the "white rot " group in that it invades xylem tissues both inter- and intracellularly and can penetrate cell walls (Spiers and Hopcroft, 1988). Miyairi and colleagues (1979, 1985, 1988) studied silver-leaf disease caused by *C. purpureum* in apple and concluded the following: 1.) The characteristic symptoms of silver-leaf disease were caused by endopolygalacturonase. 2.) This fungus was able to produce at least four endoPGs having different isoelectric points and all of these induced similar symptoms on apple leaves. 3.) EndoPG was produced in xylem by the fungus and translocated to the leaves resulting in the silvering of the leaf. 4.) Two other pectic degrading enzymes

(pectate lyase and pectinesterase) were not involved in the development of silver-leaf symptoms. 5.) There was a correlation between enzyme activity and the degree of visible disease symptoms for specific isolates of *C. purpureum*. Research at the Pacific Forestry Center has shown that endoPG is one of the major components in the culture filtrates of *C. purpureum* and that purified endoPG can cause necrosis of leaf discs. Moreover, when different isolates of the fungus were applied to the cut stumps of red alder, the extent of the disease development and plant mortality correlated well to the degree of pectic enzyme production of the fungus (Rathlef, S., unpublished data).

In *Chondrosterum purpureum*, endoPG has no doubt been shown to be involved in the disease process of the host. However, previous studies are not sufficient to answer the following questions:

1. Is endoPG required for the pathogenicity of *C. purpureum*? (The positive answer would mean that, without endoPG, the fungus ceases to be a pathogen.)

2. Does endoPG secretion correlate to the virulence of *C. purpureum*? (The positive answer would mean that, an intrinsic relationship exists between the level of endoPG secretion and the virulence level of the fungus.)

Since pathogenicity is usually a multiallelic trait, more refined approaches are needed to elucidate the precise role of one pathogenicity factor. Our rationale is to construct strains of *C. purpureum* which contain either (*i*) a disrupted endoPG gene such that no functional enzyme is produced or (*ii*) multiple copies of endoPG gene such that the enzyme will be overexpressed. It is expected that these engineered strains will have

essentially identical genetic backgrounds except for the number of endoPG genes and will differ only in their secretion of the enzyme. Pathogenicity tests of these recombinant strains together with the wild-type strain would give us answers to the above questions.

The first step in this program is the cloning and characterization of the endoPG gene from *C. purpureum*.

4.0 Summary

An endopolygalacturonase gene was identified from a genomic library of *C. purpureum*. RT-PCR experiments confirmed the processing of a single cDNA molecule from the gene and both the structural gene and the cDNA were sequenced. Sequence comparison of the genomic copy of the gene with the cDNA showed that the gene was interrupted in its coding region by nine introns. The consensus polygalacturonase active site was found in the deduced peptide sequence, showing that the gene indeed encoded a polygalacturonase. Peptide sequence comparison of the encoded polygalacturonase with fungal endoPGs and fungal exoPGs suggest that this PG belongs to the endoPG family of sequences. Southern hybridization and RT-PCR experiments indicated that there was a single copy of the endoPG encoding gene in the *C. purpureum* genome.

The cloning and characterization of the endoPG gene will facilitate the construction of the endoPG minus and the endoPG overexpressing *C. purpureum* strains. This work has laid a foundation for the refined assessment of the role endoPG plays in the pathogenicity of *C. purpureum*.

Materials and Methods

1.0 Fungal culture

Chondrostereum purpureum isolate 2128U was used in the experiments. 2128U was a monokaryotic isolate germinated from a basidiospore of the heterokaryotic isolate 2128 isolated from bigleaf maple (*Acer macrophyllus* Pursh.) in the area of Mesachie Lake of British Columbia in Canada. A starting culture was made by inoculating *C. purpureum* mycelium from 1.25% malt extract (Sigma) agar in 1.25% malt extract liquid medium. After growing at room temperature for a week, the mycelium was blended in a sterile Warring blender at low power setting for 15 seconds and was aliquoted into more flasks or petri dishes containing 1.25% malt extract liquid medium to grow at room temperature for another 1 - 2 weeks. Cultures for RNA extraction were grown in 0.1% pectin medium (Wahlstrom et al., 1991).

2.0 Nucleic acid extraction

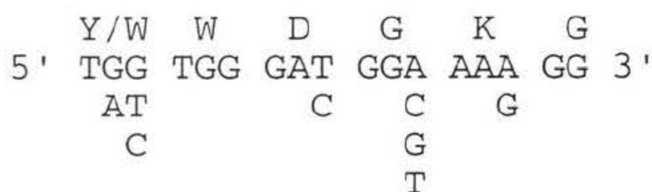
Total DNA was obtained from *C. purpureum* mycelium using the protocol for fungal DNA extraction described previously (Ramsfield, et al, 1996). Total RNA was extracted from *C. purpureum* mycelium by either the CsTFA ultracentrifugation method using the RNA extraction kit (Pharmacia) or the modified phenol-free CsTFA method using the Quick Prep total RNA extraction kit (Pharmacia). The extractions were carried out according to the manufacturer's instructions.

3.0 Amplification of an EndoPG gene fragment as a probe by RT-PCR

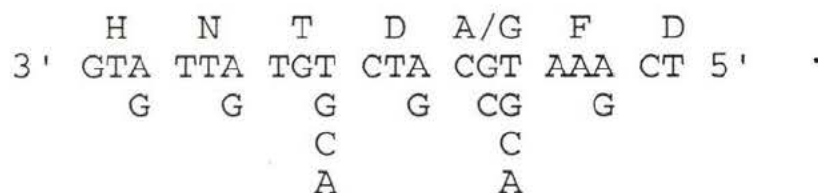
3.1 Design of the degenerate PCR primers from multiple sequence alignment

EndoPG oligonucleotide primers were designed according to conserved regions of the multiple alignment of endoPG peptide sequences of fungi, plants and bacteria (Fig. 3).

primer For-A (17nt, degeneracy 96)



primer Rev-D (20nt, degeneracy 512)



primer Rev-E (20nt, degeneracy 512)

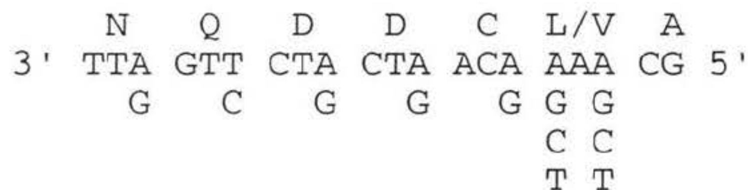


Figure 3: Design of degenerate PCR primers from multiple sequence alignment

EndoPG oligonucleotide primers were designed according to the conserved regions of peptide sequence (shown as single letter code). The peptide sequence was derived from multiple sequence alignment of endoPG sequences from fungi, plants and bacteria.

The three regions were chosen based on the conservation of their amino acids and the relatively low redundancy in the codon usage of these amino acids. These oligonucleotide primers were designed and kindly provided by Dr. P. Karlovsky (University of Hohenheim, Germany).

3.2 RT-PCR strategies

Total RNA was obtained using CsTFA ultracentrifugation method (Pharmacia RNA extraction kit). The yeild and purity of the RNA was measured spectrophotometrically (Maniatis et al, 1982). A StrataScript RT-PCR kit (Stratagene) was used. The RT-PCR procedures were outlined in Figure 4.

For the first strand cDNA synthesis, 6.8 μg of total RNA in 38 μl of DEPC-treated water was used with 3 μl of oligo(dT) primer (100 ng / μl). The first strand cDNA synthesis was done according to the instructions of the manufacturer of the kit. The subsequent PCR amplifications were carried out in a Stratagene Robocycler 40, as follows:

In step 1, two initial PCR amplifications were performed. 5 μl of the first strand cDNA reaction (1/10 of the total cDNA synthesized) was directly used as template in both. The primers were 200 pmol each of ForA with RevE (AE) and 200 pmol each of ForA with RevD respectively. The rest of the components for PCR was: 10 μl of the 10xTaq DNA polymerase buffer (Pharmacia), 0.8 μl of 25mM dNTPs plus sterile double distilled water

to give a total volume of 99.5 µl. The components were added into a sterile 500 µl microcentrifuge tube. After the careful layering of a drop of mineral oil over the reaction,

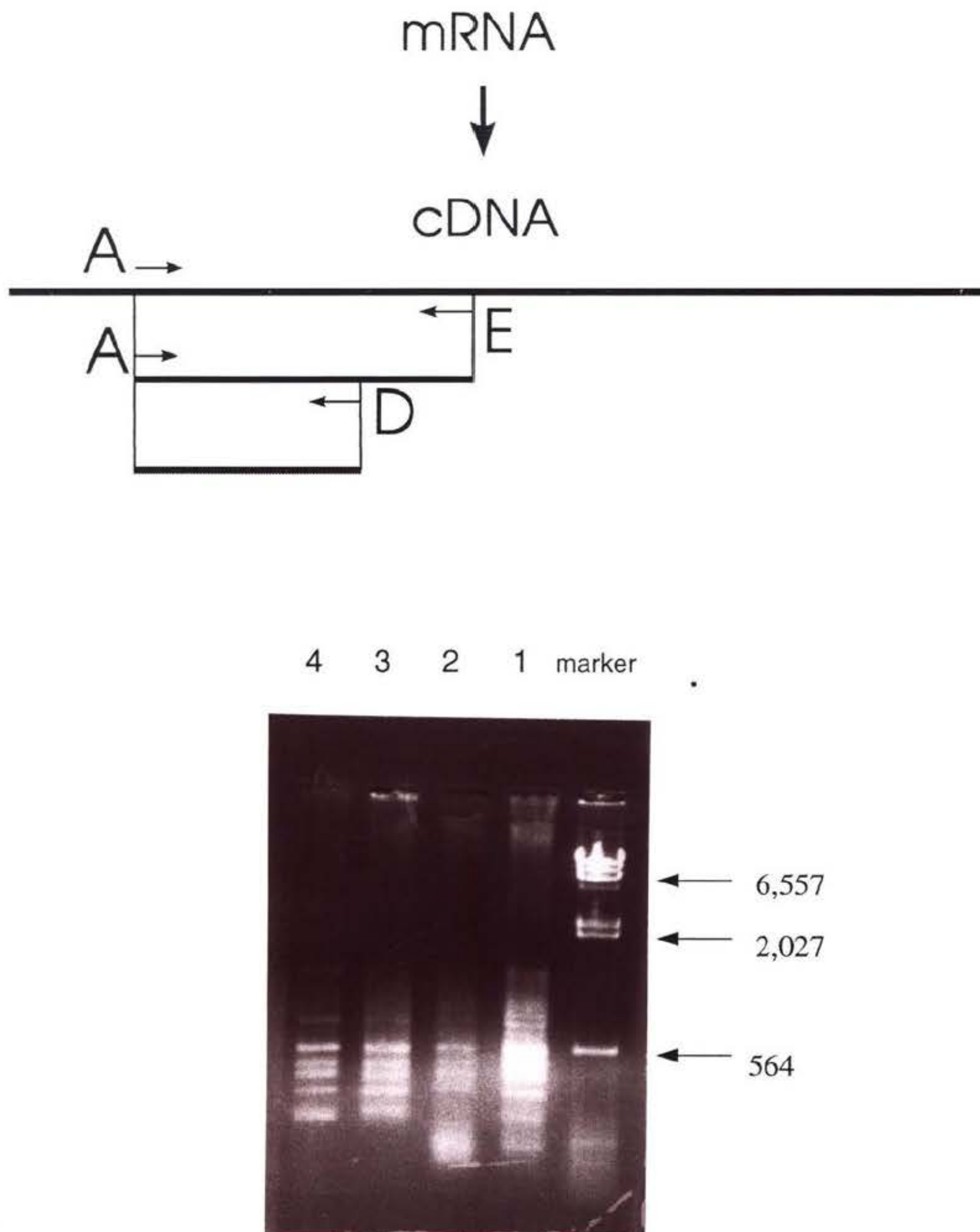


Figure 4: RT-PCR amplification of an endoPG gene fragment

Initial RT-PCR amplifications were performed with primers ForA and RevD (lane 1), and ForA and RevE (lane 2) respectively. Heminested amplifications were performed using 1/10 or 1/100 diluted products from initial RT-PCR reaction 2 as template, and using ForA and RevD as primers (lane 3 and lane 4).

the tube was put in the thermocycler at 91°C for 5 minutes and then moved immediately to 54°C for 5 minutes. A 0.5 μl aliquote (5 U / μl) of the Taq DNA polymerase (Pharmacia) was then added to the reaction. PCR amplification of the reaction was performed for 30 cycles under the following conditions: 91°C for 1 minute and 30 seconds, 54°C for 1 minute and 30 seconds, and 72°C for 2 minutes and 30 seconds. An additional 10 minutes of extension at 72°C was conducted after the completion of the 30 cycles.

Heminested PCR was performed by using 1 μl of the 1/100 diluted PCR products from the above AE reaction as template with 200 pmol each of the primers ForA and RevD. The remaining components and PCR amplification conditions were the same as described for the initial PCRs.

In all the above PCR performances, control reactions containing all the reaction components as described but excluding (i) the template DNA, (ii) both primers, (iii) each of the primers singly or (iv) the DNA polymerase were included.

3.3 Cloning and sequencing

PCR products from heminested PCR, as well as the two initial RT- PCRs were size separated on low-melting-temperature agarose (Sigma) gel. All the individual bands in the expected size range were excised, DNAs were purified by the Wizard PCR preps DNA purification system (Promega) and cloned into the pGEM-T vector (Promega). The

cloned PCR products were sequenced by dideoxynucleotide chain termination method using T7 sequencing kit (Pharmacia) with a M13 Reverse primer (Pharmacia).

4.0 Southern hybridization

Total DNA of *C. purpureum* was digested overnight with the restriction enzymes EcoRI, HincII, HindIII, HinfI, NcoI, PstI, SacI, SacII and SmaI (Pharmacia) respectively. The DNA digests were size separated on 0.7% TBE agarose gel by electrophoresis at 30V for 12 hours. The gel was then blotted onto a nylon membrane (Genescreen, Dupont).

A ^{32}P labeled endoPG gene fragment probe was synthesized using the PCR method by incorporating ^{32}P-dCTP in the amplification process (Mertz, et al., 1994).

Southern hybridization of the *C. purpureum* total DNA digests with the endoPG probe was performed in QuickHyb hybridization solution (Stratagene) as follows: 10 ml of the hybridization solution was added to a roller bottle with the membrane to incubate at 68°C for 20 minutes; the probe ($>10^8$ total counts / 1 ml of hybridization solution) together with 100µl of 10mg/ml sea urchin sperm DNA was boiled for 2 minutes and added to the preincubated hybridization solution. The hybridization was at 68°C for 2 hours. The membrane was washed twice in 2xSSC buffer and 0.1% SDS wash solution with gentle agitation at room temperature for 15 minutes each and then at 60°c with 0.1%xSSC buffer and 0.1% SDS wash solution for 30 minutes.

5.0 Genomic library construction and screening

Chondrostereum purpureum total DNA was isolated and partially digested with Sau3A so that the average size was between 10 to 20 kilobase pairs as determined by agarose gel electrophoresis. The partially digested DNA was ligated to BamHI-EcoRI predigested EMBL-3 lamda DNA replacement vector (Promega Scientific, Mississauga, Ont.) and the ligated DNA was packaged *in vitro* using commercial Lambda phage particle extracts (Gigapack, Stratagene, San Diego, CA). The *E. coli* host strain LE 392 was used to propagate the recombinant phages.

The genomic DNA library of *C. purpureum* was plated on a LE392 lawn and plaques were lifted onto nylon membranes (Genescreen, Dupont). Plaque hybridization was done according to Maniatis et al (1982), using the ^{32}P labeled endoPG fragment as a probe. The positive phage clones were picked , eluted, replated and plaque hybridization sequentially repeated twice to obtain well-isolated, positively hybridizing phage clones. Lambda phage DNA from a positive clone, clone7 was purified using purification of Lambda DNA with Wizard Lambda preps DNA purification system from Promega.

6.0 Subcloning

The endoPG lambda DNA was digested with SalI, EcoRI, BglII and BamHI respectively. The digested lambda DNA fragments were size separated on agarose gel by electrophoresis, blotted onto a nylon membrane (Genescreen, Dupont) and hybridized with the ^{32}P labelled endoPG fragment to determine the best sub-cloning strategy. Both

the whole digested lambda DNA mix and the DNA purified from the positively hybridizing individual restriction fragments excised from gels were ligated into appropriately cut plasmid vector PUC18 (Pharmacia) or PTZ18R (Pharmacia) (Fig. 5).

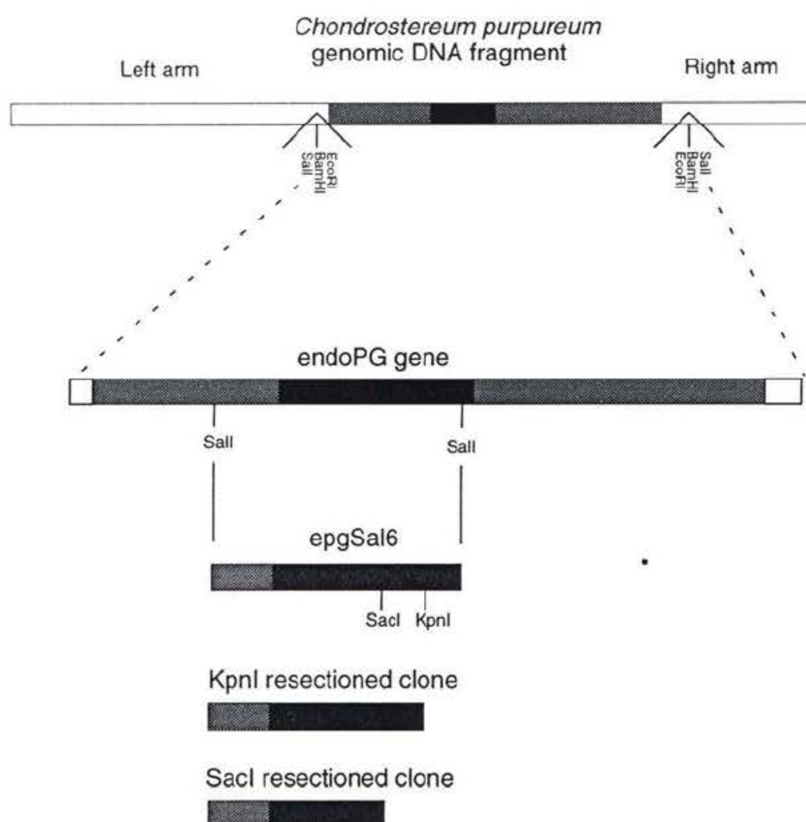


Figure 5: Gene cloning outline

The genomic DNA of *C. purpureum* was partially digested with endonuclease Sau3A and cloned onto lambda DNA replacement vector EMBL-3. The clone which contained endoPG gene was retrieved using a RT-PCR derived gene fragment as a probe. A SalI fragment of 3.28 kb from this phage DNA, which contained most of the gene was further subcloned onto a plasmid vector PUC18 and was termed ePGsal6. To facilitate the sequencing of the gene, two endonucleases KpnI and SacI, both having a unique site inside the ePGsal6 clone, were used individually to cut the clone and the digested clone was recircularized and termed KpnI resectioned clone and SacI resectioned clone respectively.

Two methods were used to screen the subclones for the presence of the endoPG gene fragment amplified by heminested PCR. One was to blot the digested DNA of the subclones and to hybridize with the ^{32}P labelled endoPG fragment. The other was PCR screening using the plasmid DNA from the subclones as template with the degenerate primer ForA (Fig.3) and a nondegenerate primer ePGcon1 (5'TATTTCC GATAGATATTGCCTGCGC3'). The positive control reaction used the cloned endoPG gene fragment from RT-PCR as template with primers ForA and ePGcon1, in which a 130bp DNA was always amplified. Each test template DNA was constructed as a mixture of four or eight plasmid DNAs so that if a 130bp DNA was amplified, the mixed template DNA would be split into individual plasmid DNA and each be used as a template in PCR to identify the positive clone.

PCR experiments to determine the orientation of the insert and to estimate how much of the endoPG gene was in the endoPG subclone pePGsal6 were carried out using vector-specific primers (Universal forward primer and M13 reverse primer, Pharmacia) and endoPG specific primers (ePGcon1 and ePGcon2: 5'CGGAGATAGTGGTAGTCTCG GGCAC3') (Fig. 6).

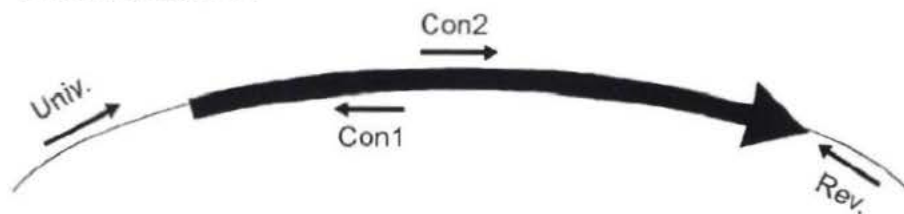
7.0 Sequencing strategies for the genomic copy of the endoPG gene

Plasmid DNA of pePGsal6 was sequenced by the dideoxynucleotide method according to the manufacturer's instructions using the T7 sequencing kit from Pharmacia. To obtain

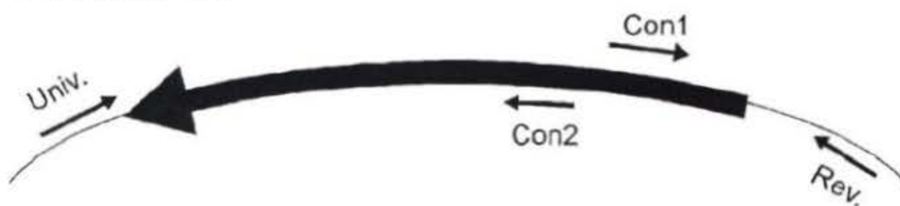
Figure 6: PCR used to determine the insert orientation in the subclone pePGsal6 and to estimate how much of the gene was contained in the clone.

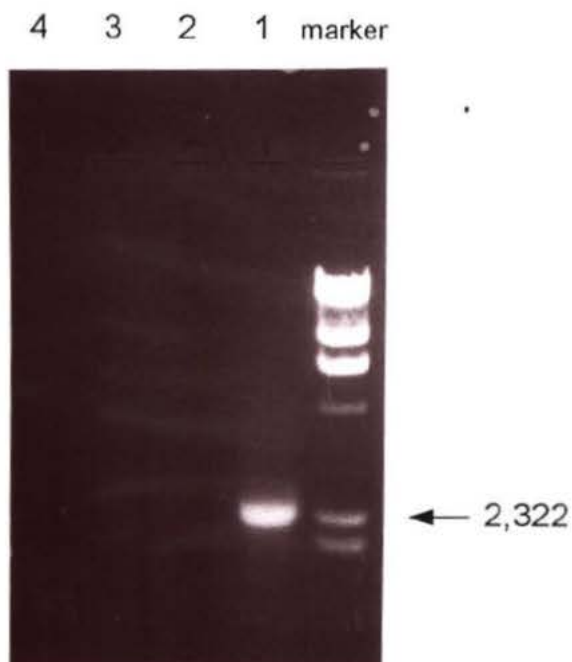
There were two possible orientations concerning the cloned endoPG gene containing SalI fragment in the plasmid subclone ePGsal6. Estimation of how much of the entire gene was contained on the plasmid was useful in deciding if this plasmid was sufficient for the sequencing of the gene. After two endoPG gene fragment containing subclones were identified, the smaller clone, ePGsal6 was subjected to the above PCR analysis before proceeding to sequencing it. Four primer combinations were used individually with ePGsal6 DNA as the template DNA in PCR. Primer combination 1 amplified a fragment of about 1.2 kb, suggesting that the orientation of the gene contained in ePGsal6 was the same as the vector PUC18. By knowing the length of the PCR amplified gene fragment and the location corresponding to its 3' end on the consensus endoPG peptide sequence derived from alignment, it was estimated that most of the coding sequence of the gene was contained in ePGsal6.

Orientation A



Orientation B





Primer combinations used in the four PCRs:

A: 1. Univ. + Con1 (lane 1) 2. Con2 + Rev. (lane 2)

B: 3. Univ. + Con2 (lane 3) 4. Con1 + Rev. (lane 4)

more sequence information using M13 reverse primer, KpnI and SacI resectioned clones were made from pePGsal6 by using KpnI and SacI respectively to delete either a KpnI or a SacI fragment from the clone and then self-ligating the DNA (Fig. 5). Sequencing reactions were first performed using universal forward primer, M13 reverse primer and primers corresponding to the PCR amplified endoPG gene fragment region inside the clone. Sequencing reactions were also done with the KpnI and SacI resectioned clones using M13 reverse primer. From the sequences obtained, new primers were designed and used for further sequencing (Fig. 7). Primer design was done with the computer program OLIGO to avoid formation of primer dimers or internal hairpin structures or runs of one type of nucleotide, especially G/C, at the 3' end. Computer analysis of the sequence information was performed using the program Generunner (Hastings Software Inc.,1994) for the following: compilation of the entire sequence, prediction of the reading frame, prediction of the secondary structure (Chou and Fasman, 1978; Garnier, Osguthorpe, and Robson, 1978) and the hydrophobicity of the deduced peptide (Kyte and Doolittle, 1982) as well as the potential N-glycosylation sites in the peptide.

The sequence at the 3' end of the coding region of the endoPG gene was not contained in the subclone pePGsal6 hence had to be obtained by direct sequencing of the endoPG lambda DNA using the CircumVent thermal cycle dideoxy DNA sequencing kit from New England Biolab. The primer used was designed from the 3 prime region of the incomplete endoPG gene in pePGsal6 and was ^{32}P end labelled using [γ-^{32}P] rATP and T4 polynuleotide kinase according to the procedures on the instruction manual of the sequencing kit. A Stratagene Robocycler 40 was used to perform the sequencing

reactions and conditions for the reactions were as follows: 95°C, 1 minute; 50°C, 1 minute and 72°C, 1 minute, for a total of 20 cycles.

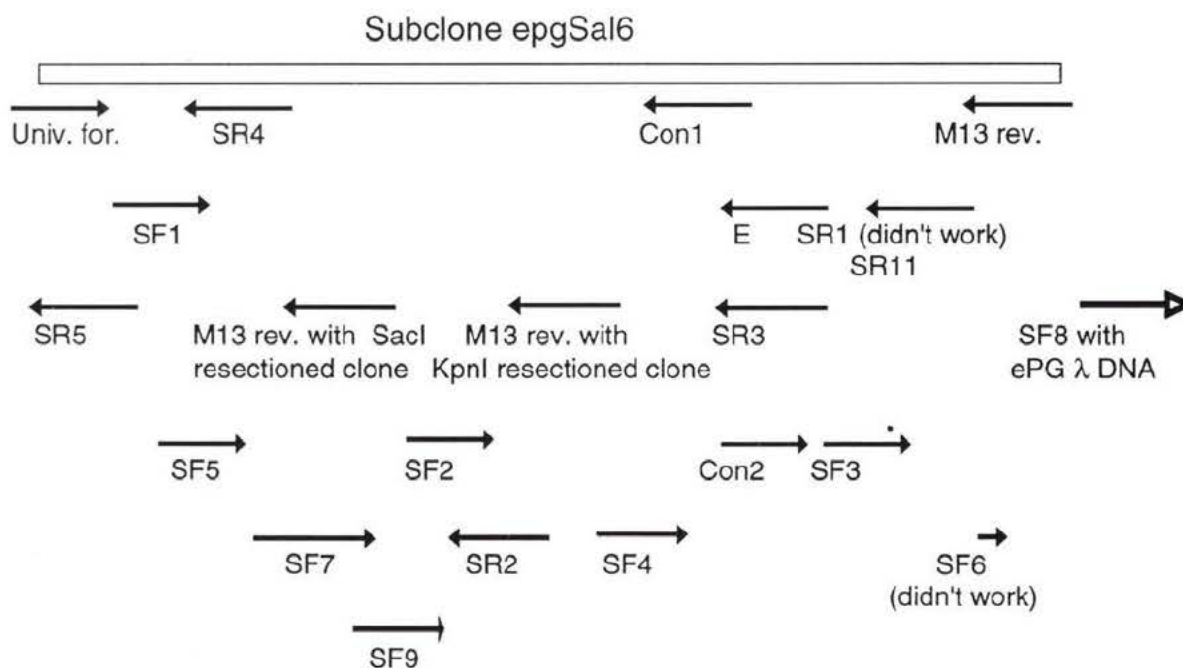


Figure 7: Sequencing strategy

Custom designed primers together with universal forward primer and M13 reverse primer were used with ePGsal6 plasmid DNA in T7 sequencing reactions. M13 reverse primer was also used with both KpnI resectioned and SacI resectioned clones in T7 sequencing. The very 3' end of the gene, around 280 bps, which was not in ePGsal6 clone, was derived from direct sequencing with the ePG lambda DNA using ^{32}P end labeled primer SF8. Nearly the entire sequence of the gene was obtained from both strands.

8.0 cDNA amplification by RT-PCR

8.1 Primer design

Three primer combinations (Fig. 8) were used for the RT-PCR amplification of three overlapping endoPG cDNA fragments, endoPG cDNA fragment 1, cDNA fragment 2 and cDNA fragment 3, which together make up the endoPG coding region. For the amplification of the endoPG cDNA fragment 1 the forward primer ePGcFor2 (5'CCAT GCCTTCGTTATCCTCGATCCT3') was designed from the putative translation initiation site of the endoPG gene and the reverse primer ePGcRevD (5'ACATCGAAGCCAT CAGTGTTGTGCC3') was designed from a short region spanning an intron (identified by comparing the sequence of the short endoPG gene fragment derived by RT-PCR using the degenerate primers with the sequence of the genomic copy of the gene) hence would only anneal properly to fully processed cDNA. For the amplification of the endoPG cDNA fragment 2, the forward primer ePGcForD (5'CACTGATGGCTTCGATGTATCT3') was designed from the same region as that for the primer ePGcRevD (hence was also a cDNA specific primer) and the reverse primer was cRev2 (5'GTCAGCTTAGACCAGTT CCACG3'). For the amplification of the cDNA 3, the forward primer cFor3 was designed from the same region as cRev2, and the reverse primer was a GC-rich-tailed poly$(dT)_{18}$ (5'GGCGAGGTTTTTTTTTTTTTTTTTT3').

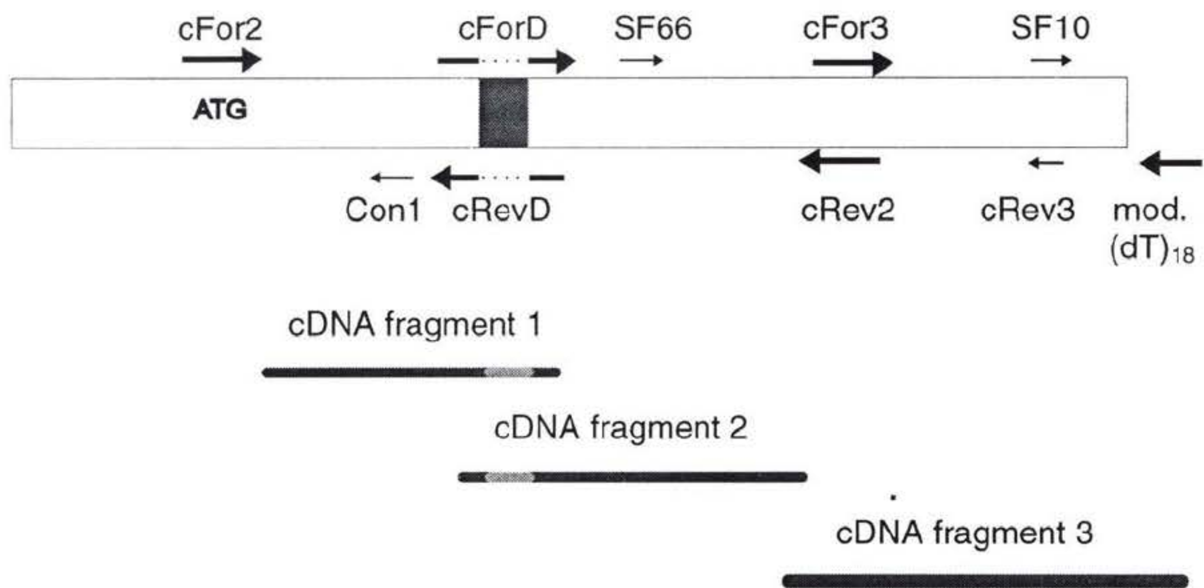


Figure 8: cDNA amplification by RT-PCR

Outline of primer combinations used in the RT-PCR amplification of overlapping cDNA fragments of the polygalacturonase gene from *Chondrostereum purpureum*. Note: An intron is indicated within cForD/cRevD region.

8.2 RT-PCR strategies

Total RNA used here was isolated according to the manufacturer's instructions using the modified phenol-free CsTFA method. The RNA obtained was quantified by gel electrophoresis. The first strand cDNA was synthesized using approximately 1 μg of total RNA in 24 μl of DEPC-treated water with 1 μl (0.5 μg) of the oligo(dT), according to the protocol for first strand cDNA synthesis of transcripts with high GC content, using the SuperScript preamplification system from GibcoBRL. A 5 μl aliquot of the first strand cDNA reaction (1/10 of the total reaction) was used directly as the template in subsequent PCR amplifications.

For the PCR amplification of endoPG cDNA fragment 1, 80 pmols each of the primers ePGcFor2 and ePGcRevD, 10 μl of the 10xPCR buffer (200mM Tris-Hcl, PH 8.0, 500mM Kcl), 6 μl of 25mM Mgcl, 2 μl of 10mM dNTP mix, 1 μl of Taq DNA polymerase (5 U / μl, Pharmacia) were used together with sterile double distilled water to a total reaction volume of 100 μl. The conditions for the PCR amplifications were as the following: 94°C for 3 minutes to denature the RNA/DNA hybrid, followed by 5 cycles of 94°C for 1 minute and 30 seconds, 60°C for 1 minute and 30 seconds and 72°C for 2 minutes and 30 seconds, followed by 25 cycles of 94°C for 1 minute and 30 seconds, 65°C for 1 minute and 30 seconds and 72°C for 2 minutes and 30 seconds. A final extension at 72°C for 10 minutes was performed after the 30 cycles. After this, a heminested PCR was performed, using 1 μl of the above reaction as template with 80 pmols each of primers ePGcFor2 and ePGcon1 (Fig.8). The rest of the components and

the amplification conditions for this PCR were the same as described in the above, except that the first 5 cycles were omitted.

For the PCR amplification of the endoPG cDNA fragment 2, 5 μl of the cDNA was used as template together with 80 pmols each of primer cForD and primer cRev2, 10 μl of 10xTaq buffer, 2 μl of 10mM dNTPs, 1 μl (5U) of the Taq polymerase and double distilled water to a total reaction volumn of 100 μl. The reaction was placed at 94°C for 3 minutes to denature DNA/RNA hybrid and then subjected to the following cycle conditions: 94°C for 1:30 minutes, 50°C for 1:30 minutes and 72°C for 2:30 minutes for a total of 30 cycles. The sample was then placed at 72°C for another 10 minutes to polish the products. A heminested PCR was performed using 1μl of the above PCR product as template and 80 pmols each of primer SF66 and cRev2 (Fig.8) with the rest of the reaction components the same as described above. The cycle conditions were as follows: 94°C for 1:30 minutes, 55°C for 1:30 minutes and 72°C for 2:30 minutes, for a total of 30 cycles. A reamplification of the original PCR products was also conducted.

The primer cFor3 and the modified oligo$(dT)_{18}$ were used for the amplification of cDNA fragment 3. Because the modified oligo(dT) primer was not a specific primer for cDNA fragment 3, only the forward primer was included during the first 4 cycles of PCR to generate some specific single strand DNA which could serve as template in the subsequent cycles. The initial PCR conditions included: 5 μl of the cDNA mix as template DNA with 2 μl of 10mM dNTP, 10 μl of 10x Taq buffer, 5 units of Taq DNA polymerase, 43 pmols of primer cFor3 and sterile double distilled water to a total reaction volumn of 99 μl. This sample was run under the following conditions: 94°C for 3

minutes to denature the DNA/RNA template, then at 94°C for 1:30 minutes, 55°C for 1:30 minutes and 72°C for 2:30 minutes for 4 cycles. The modified oligo$(dT)_{18}$ primer (47 pmol in 1µl) was added to the sample after the first 4 cycles and the sample was run for another 35 cycles under the same conditions. A final 10 minute incubation at 72°C was added to polish the products. Heminested PCR and reamplification of the above PCR was also performed. For heminested PCR, 2 µl of the above PCR product was used as template with the rest of the components under identical conditions except that the primers used were either primer cFor3 (42 pmols) with primer cRev3 (43 pmols) or primer SF10 (43 pmols) with modified oligo(dT) (47 pmols) (Fig.8). The amplification conditions for the heminested PCR were as follows: 94°C for 1:30 minutes, 55°C for 1:30 minutes, 72°C for 2:30 minutes for a total of 30 cycles.

8.3. Sequencing cDNA fragments

Two different approaches were used in sequencing the three overlapping cDNA fragments derived by RT-PCR, based on the differences in the RT-PCR amplification results.

For cDNA fragment 1 and cDNA fragment 2 where RT-PCR yielded distinctive individual DNA fragments, the PCR products from the heminested PCR (cDNA fragment 1) or the reamplified original PCR (cDNA fragment 2) were size separated on low-melting-temperature agarose gel. The band of the expected size was excised, DNA was purified from the excised band by the Wizard PCR preps DNA purification system (Promega) and cloned into the pGEM-T vector (Promega). The cloned PCR products

were sequenced by dideoxynucleotide method using T7 sequencing kit (Pharmacia) with a M13 Reverse primer and a universal forward primer (Pharmacia).

For cDNA fragment 3 where the products of RT-PCR was more of a smear than distinctive bands when visulized under UV light after the products were subjected to electrophoresis and EB staining, direct sequencing with the whole PCR products using specific internal forward and reverse primers was performed. The specific primers, cRev3 and SF10, were designed from a region judged to be exon and situated in the middle part of the cDNA fragment 3 (Fig. 8). Excess primers were removed from reamplified cDNA fragment 3 RT-PCR by direct column purification using Wizard PCR preps purification kit (Promega) and redissoved in 40 μl of double distilled water. 8.5 μl of the DNA each was used as template DNA with 1.2 pmol of primer cRev3 or SF10 respectively in thermal cycle sequencing reaction with the incorporation of 35S-dATP (New England Biolab). The reaction was carried out according to the recommendation of the manufacturer except for the cycle reaction conditions which was as follows: 94°C for 1 minute, 58°C for 1 minute and 72°C for 1 minute for a total of 20 cycles.

9.0 Phylogenetic analysis using polygalacturonase

The deduced endoPG peptide was used in BLAST search for similar protein sequences using all the protein data bases and the translated DNA data bases. All the full-length polygalacturonases thus retrieved were used in phylogenetic analysis except in situations when there were several proteins from the same organism which varied only by several amino acids. In these cases one representative protein was used and the others were

excluded from the study. There were altogether 34 polygalacturonases used in the study: two of them from bacteria, sixteen from fungi and sixteen from plants. The phylogenetic analysis was done using progarms belonging to HUSAR (Heidelberg Unix Sequence Analysis Resources) suite Version 4.0, installed on Convex Exemplar SPP-1200 in the German Cancer Research Center (Heidelberg) and available to German universities through GENIUSnet. The sequences were aligned by PILEUP(+). After the alignment, a distance matrix was generated using the program DISTANCES and a phylogenetic tree was reconstructed by GrowTree.

The multiple alignment of 34 protein sequences was constructed using PILEUP(+). The sequences were first clustered using similarities calculated for each pair. The similarities were used to construct a preliminary dendrogram using UPGMA (Sneath and Sokal, 1973). The dendrogram directed the order of the subsequent alignments. The multiple alignment procedure began with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. Each pairwise alignment in PILEUP used the method of Needleman and Wunsch (1970), which was extended for use with clusters of aligned sequences rather than only individual sequences. This cluster was then aligned to the next most related sequence. The final alignment was achieved by a series of progressive, pairwise alignments that included increasingly dissimilar sequences and clusters, until all sequences had been included. For a pairwise alignment of clusters of sequences, the comparison score between any two positions in those clusters was simply the arithmetic average of the scores for all possible symbol comparisons at those positions. When gaps were inserted into a cluster to produce an alignment, they were inserted at the same position in all of the sequences of the cluster.

The parameters used for the alignment using PILEUP(+) were: Gap creation penalty: 3.00; Gap extension penalty: 0.10; Symbol comparison table: pileuppep.cmp.

After the alignment, a distance calculation was performed using the program DISTANCES. DSTANCES wrote a matrix of the pairwise evolutionary distances between aligned sequences. The distances were expressed as substitutions per 100 bases or amino acids. Each pair of aligned sequences was examined symbol-by-symbol and the number of exact matches, partial matches, and gap symbols were counted. Several methods may be chosen to correct the distances for multiple substitutions at a site. In this work, Kimura's correction for protein sequences was used (Kimura, 1983) with Dayhoff's matrix. This formula calculated distances based on the relationship between observed amino acid substitutions and actual (corrected) substitutions that was derived by Dayhoff and coworkers (1979). Gap positions were ignored, and only exact matches contributed to the match score.

Program GrowTree was used to reconstruct a phylogenetic tree from the distance matrix created by DISTANCES. Two methods were available for reconstructing the tree: UPGMA and neighbor-joining. The second method was used in this study. Neighbor-joining method was designed to find an approximation to the minimum evolution tree for a set of aligned sequences. It clustered the sequences in a pairwise fashion. Instead of picking the next pair to cluster by looking for the smallest distance in the distance matrix, the method sought to form pairs that minimize the sum of the branch lengths for the entire tree. Therefore at each round of clustering, all possible pairs of entries were considered one at a time and the sum of the branch lengths for the resulting tree was

calculated. The pairing that resulted in the smallest sum was the one that used to form the new cluster. This new cluster replaced its two constituent entries in the distance matrix (reducing the dimension of the distance matrix by one), and distances were calculated between the new cluster and the remaining entries in the distance matrix. The process was continued until only two entries remained. The method was designed by Saitou and Nei (1987) and simplified by Studier and Keppler (1988). The following were the parameters used for the reconstruction of the phylogenetic tree: -NONEGative: resets negative branch lengths to zero; -ORDer=0: ordering of sequences in tree display; -TREEFORMat=1: format to use in drawing the tree such that branch lengths are proportional to calculated distances; -NOPLOt: suppresses graphical display of tree; -FIGure[=FileName] stores plot in a file for printing on a Postscript printer instead of plotting it on a HP-Plotter.

Results

1.0 Cloning of the endoPG gene from *C.pupureum* genomic library

1.1 PCR cloning of a fragment of endopolygalacturonase gene

In order to retrieve the endoPG gene from the genomic DNA library of *Chondrostereum purpureum*, a suitable DNA probe had to be derived. Several methods were attempted.

(1) Heterologous probe

A cloned endoPG cDNA from the ascomycete fungus *Fusarium moniliforme* obtained from Dr. F. Cervone (Caprari, et al., 1993) was radio labelled with ^{32}P dCTP by random priming method to use as a hybridization probe against the genomic DNA digests of *C. purpureum*. The hybridization was repeated twice with decreasing stringency in washing but each time no hybridization signal was detected. This is not surprising in view of the overall low homology among endoPG gene sequences from different species.

(2)PCR cloning

Several approaches were taken to amplify a fragment of endoPG encoding gene by means of the Polymerase Chain Reaction (PCR) using synthesized oligonucleotides as primers and genomic DNA/cDNA of *C. purpureum* as a template. It would have been ideal to have some authentic amino acid sequence of endoPG of *C. purpuruem* to aid in primer design but the endoPG enzyme had not yet been purified from *C. purpureum*. We made use of published endoPG gene sequences from other species including bacteria, plants and fungi to design PCR primers. Comparison of the derived amino acid sequence of

endoPG from several species showed an overall low level homology but several relatively conserved short regions existed. Oligonucleotide primers, corresponding to these regions, were designed. Initially, the primers were designed using the codon preference of the ascomycete fungus *Aspergillus nidulans*. Several rounds of PCR amplification using different combination of these primers followed by cloning and sequencing of the PCR products, and comparison of these sequences to published endoPG sequences showed no similarity of the PCR products to the endoPG genes. Because *C. purpureum* is basidiomycete, and all of the published fungal endoPG sequences were from ascomycetes, it was possible that this endoPG was very different. Southern hybridization was performed using each cloned PCR product as a probe to hybridize with genomic DNA digests of *C. purpureum*. Again, no hybridization signals were detected. In collaboration with Dr. P. Karlovsky (University of Hohenheim, Germany), three degenerate primers were designed according to three well-conserved short regions from multiple endoPG peptide sequence alignment (Fig. 3).

At first, the degenerate primers were used with genomic DNA of *Chondrostereum purpureum* in PCR amplification but nothing was amplified. A similar result was reported for the cloning of endoPG gene of *Fusarium moniliforme*. This was later found to be due to one of the primers being split by an intron in genomic DNA. We conjectured that this could well have been the reason for our PCR failure here, hence cDNA was used as template instead of the genomic DNA. Heminested RT-PCR was performed to amplify a fragment of the endoPG gene from *C. purpureum*. PCR products from hemi-nested PCR, as well as the initial RT-PCR were size separated on agarose gel. Hemi-nested PCR yielded one predominant band among several weak bands in the expected

size range while the initial RT-PCRs yielded several weak bands in their respective expected size ranges (Fig. 4). No amplification was detected with the control PCRs. All the individual bands in the expected size range were excised and DNAs were purified and cloned. The cloned PCR products were sequenced and derived peptide sequences of the PCR products were compared with endoPG peptide sequences from other organisms. From the hemi-nested PCR, a 214bp product whose translated peptide sequence revealed quite significant similarity with other endoPGs, was obtained (Fig. 9). This was termed the endoPG gene fragment. This fragment was synthesized and ^{32}P labeled by the incorporation of $[\alpha\text{-}^{32}P]$dCTP in the PCR for use as a hybridization probe (Mertz, et al., 1994). When it was used to hybridize with genomic DNA digests of *C. purpureum,* a single band was detected in each lane of the DNA digests showing that this fragment was amplified from a single-copy gene in *C. purpureum* genome (Fig. 10).

```
                                           Primer ForA ===>
C. purpureum         YWDGQGTNG-GVTKPHPF-LKFKGSGQYSSFT
S. sclerotioru       ASGSIL ----DGNGGKYWDGKGTNG-GKTKPK-FFYAHSLKGKS-SIN
C. carbonum          ASGHTI ----DAAGQKWWDGKGSNG-GKTKPK-FFYAHSLTT-S-SIS
F. moniliforme       ASGHVI ----DGNGQAYWDGKGSNSNSNQKPDHFIVVQKTTGNS-KIT
A. oryzae            ASGAKI ----DGDGSRWWDGKGGNG-GKPKTK-FFYAHKLDS-S-SIT
A. nidulans          AAA-VI ----DVQGSRWWDGKGPNG-GKTKPK-FIQYPQLES-P-TIT
Avocado              EGGGTI ----NGNGETWWDSSCKRKKSLPCKSAPTALTFRSCKNLIVS
Prunus persica       -SGGIL ----DGQGTALW--ACKACHGESGPSGATTLGFSDSNNIVVS
P. solanacearu       RGGSLV TSGPNANRLTWWDIAYLNKTKGLNQQNPRLIQTYNGSAFTLY
E. carotowora        QGGVKL ----QDKKVSWWDLAADAKVKKLKQNTPRLIQINKSKNFTLY
```

```
C. purpureum         ---VLNSPAQAISIGNSDGLTFDTVTVDNS----AGDS-----GSL---
S. sclerotioru       NVKILNSPEEGFSINSASGLTLSGITIDNS----GGNSL----------
C. carbonum          GLNIKNTPVQAFSINGVTGLTLDRITIDNS----AGDSA----------
F. moliniforme       NLNIQNWPVHCFDITGSSQLTISGLILDNR----AGDKPNAKSGSLP--
A. oryzae            GLQIYNTPVQGFSIQS-DNLNITDVTIDNS----AGTAE----------
A. nidulans          GLHVKNSPVQVFSVQG-NDVHLTDITIDNS----DGDNN----------
Avocado              DLSIKDSQKMHLSFDKCQDVIASNLMVTAP-------------------
Prunus persica       GLASLNSQMFHIVINDFQNVQMQGVRVSRS-------------------
P. solanacearu       GVTVQNSPNFHIVTTGTSGVTAWGIKIVTPSLAYAVAGYKCPSGSTPDK
E. carotowora        NVSLINSPNFHVVFSDGDGFTAWKTTIKTPS------------------
```

```
C. purpureum         ------------GHNTDGFD
S. sclerotioru       ------------GHNTDAFDVGSS
C. carbonum          -----------GAHNTDAFDIGSS
F. moliniforme       -----------AAHNTDGFDIGSS
A. oryzae            ------------GHNTDAFDVGSS
A. nidulans          -----------GGHNTDAFDVSES
Avocado              ----------EHSPNTDGIHITGT
Prunus persica       ----------GNSPNTDGIHVQMS
P. solanacearu       VTPATCFTPETVKNTDGFDPGQS
E. carotowora        -----------TARNTDGIDPMSS
                           <=== Primer RevD
```

Figure 9: Alignment of the peptide sequence of the putative endoPG gene fragment from *C. purpureum* with endoPGs from fungi, plants and bacteria

Figure 10: Gene copy analysis by Southern hybridization

Total DNA of *C. purpureum* digested individually with different restriction enzymes, EcoRI (lane 1), HincII (lane 2), HindIII (lane 3), HinfI (lane 4), NcoI (lane 5), PstI (lane 6), SacI (lane 7), SacII (lane 8) and SmaI (lane 9), were blotted onto a membrane and hybridized with ^{32}P labeled putative endoPG gene fragment. This fragment is shown to be amplified from a single-copy gene in the *C. purpureum* genome.

1.2 Genomic library construction and screening

The integrity of the genomic library was first checked by the recovery of the ribosomal DNA (rDNA) repeat of *C. purpureum* by screening with a heterologous rDNA probe. The derived rDNA was used in studying the geographic variation of *C. purpurem* by looking at the polymorphisms in the ribosomal DNA (Ramsfield, et al., 1996).

The ^{32}P labeled endoPG fragment was used as a probe to screen the EMBL-3 genomic DNA library of *C. purpureum.* Of 3500 phage clones screened one positive clone, endoPG lambda DNA clone 7, was identified.

The lambda DNA was digested with SalI, EcoRI, BglII and BamHI respectively. The Southern hybridization of the digested lambda DNA with the ^{32}P labeled endoPG probe showed that there was a single positively hybridizing fragment in each of these lambda DNA restriction digestions. The digested endoPG lambda DNA fragments were subcloned into PUC18 (Fig. 5). About 140 of such subclones were screened by Southern hybridization or PCR, and only two SalI clones, ePGsal2 and ePGsal6, were identified which contained the section of the endoPG gene corresponding to the amplified probe. The plasmid ePGsal6 contained a positively hybridizing Sal1 fragment between 3kb and 4kb and was chosen for further analysis.

Using combinations of vector-specific and endoPG gene specific primers, it was determined that the insert included much of the 5' untranslated region and the whole or nearly the whole coding region of the endoPG gene. Later it was found out that the very

3' end of the coding region was not subcloned onto pePGsal6. This remaining 3' region of the coding sequence however was resident on the EMBL-3 clone (see below).

1.3 Sequencing of the genomic copy of the endoPG gene

Two different approaches were used for sequencing the plasmid ePGsal6.

In DNA sequencing, the deletion of specific regions of DNA has become a convenient strategy for sequencing large DNA fragments. A popular approach is to generate a library of nested deletions representing a series of subclones with progressively more of the target sequence deleted from a common original clone. This is made possible by the controlled digestion of the DNA with exonuclease III, a 3'-exonuclease active only on double-stranded DNA, and S1 nuclease to remove the single strand left by 3'-exonuclease digestion. This method enables the determination of the entire sequence of a large cloned fragment using a single sequencing primer (for instance, the universal M13 primer). Because of the attractiveness of using the universal primer for sequencing the entire clone, nested deletions were the preferred approach for sequencing pePGsal6. Restriction analysis of the plasmid was done to determine useful sites for exonuclease digestion. Six enzymes in the multiple cloning site of the vector PUC18 were used to digest pePGsal6. EcoRI and XbaI seemed to be the only enzymes that cut the clone once.

The EcoRI site was used as filling-in B site (3'-exonuclease resistant) while XbaI site was used as A site (3'-exonuclease susceptible) in the nested deletion. The deletion was conducted according to the protocol by the supplier of the kit (Pharmacia). The whole

procedure of generating gradually deleted clones was repeated five times with no success. Hence, this approach was abandoned and direct sequencing of the clone with custom-designed primers was used.

T7 Sequencing reactions were performed with pePGsal6. Most of the sequence obtained were from both strands. When areas of sequence uncertainty arouse, they were sequenced again using either the same primer or a different one. Of all the 16 sequencing primers designed, 2 primers failed to work although they were also carefully checked to avoid formation of primer dimers or internal hairpin structures or runs of one type of nucleotide, especially G/C, at the 3' end. Occassionally the phenomenon of sequence compression was encountered during the sequencing. This was observed as of an unambiguous reading obtained from one sequencing gel film whereas when the sequencing reactions were run in the opposite direction in the same region there appeared to be a loss of one nucleotide somewhere in the sequence. The deduced amino acid compared to the consensus sequence from endoPG alignment also supported the fact that indeed in one sequencing gel that nucleotide was lost - compressed with the neighboring nucleotide.

Because of the overall low level homology among endoPG peptide sequences, and because the first custom-designed sequencing primer to the 3' end of the ePGsal6 clone did not work, it was near the completion of the sequencing work when it was found out that a very short piece of the 3' end of the coding region of the endoPG gene was missing from the SalI fragment cloned in pePGsal6. Direct sequencing of the endoPG lambda clone with primer SF8, designed from the very 3' end of the gene fragment in pePGSal6,

was performed and the sequence obtained showed that the it was the authentic 3' end of the gene. The full-length endoPG gene was contained on the lambda clone and the sequence of the gene was made complete by the cycle sequencing with the lambda DNA.

A total of 3.28 kb sequence information was obtained for the endoPG structural gene, which consisted of the coding region flanked by approximately 1.5 kb of the 5' upstream region and approximately 100 bp of the terminator region of the gene. The coding region was disrupted many times in its reading frame, suggesting the existence of 8-10 introns.

From the comparison of the genomic sequence with the sequence of the endoPG gene fragment derived by RT-PCR with degenerate primers, it was clear that the primer RevD region in the gene was split by an intron and hence it did not work when the genomic DNA was used as the template in PCR with the degenerate primers (as conjectured earlier).

2.0 Amplification of the cDNA fragments of the endoPG gene by RT-PCR

Tentative assignment of intron positions by searching for consensus 5' and 3' splice sites (GT---AG) in conjunction with seaching for switching of reading frame and the presence of extra sequence embedded within a conserved region of the endoPG gene was not possible for the entire coding region because 1.) there was an overall low level homology among all the endoPG peptide sequences, especially at regions near the two termini; 2.) the reading frame shifted many times suggesting the existence of 8 to 10 introns which is in contrast to the number of introns found in the ascomycete endoPG genes (0 to

4 introns only). Fungal genes generally have a relatively small number of exons (fewer than 6) (Lewin, B., 1994). To better characterize the cloned endoPG gene, verification of the intron/exon position was necessary. RT-PCR was therefore used to amplify cDNA fragments of the endoPG gene with the result that three overlapping endoPG cDNA fragments were obtained: endoPG cDNA fragment 1, endoPG cDNA fragment 2 and endoPG cDNA fragment 3 (Fig. 11).

2.1 Amplification of the endoPG cDNA fagment 1

Three regions in the gene sequence were considered, the translation start site, a middle region and the end of the gene, for primer design. The forward primer (ePGcFor2) was designed out of a putative translation start site judged by the sequence context of the ATG start codon. The design of the primers from the middle region of the gene made use of the intron-containing region where the degenerate primer RevD located. Specific primers ePGcForD and ePGcRevD were designed from the cDNA sequence spanning this intron (Fig.8). In this way, the two primers would anneal selectively to the cDNA only because the corresponding area in the genomic DNA was interrupted by the intron. A reverse primer (ePGcREV) was designed from the very end of the gene sequence. Three possible primer combinations from the above primers were used in RT-PCR and only primers ePGcFor2 with ePGcRevD amplified one single DNA fragment of the predicted

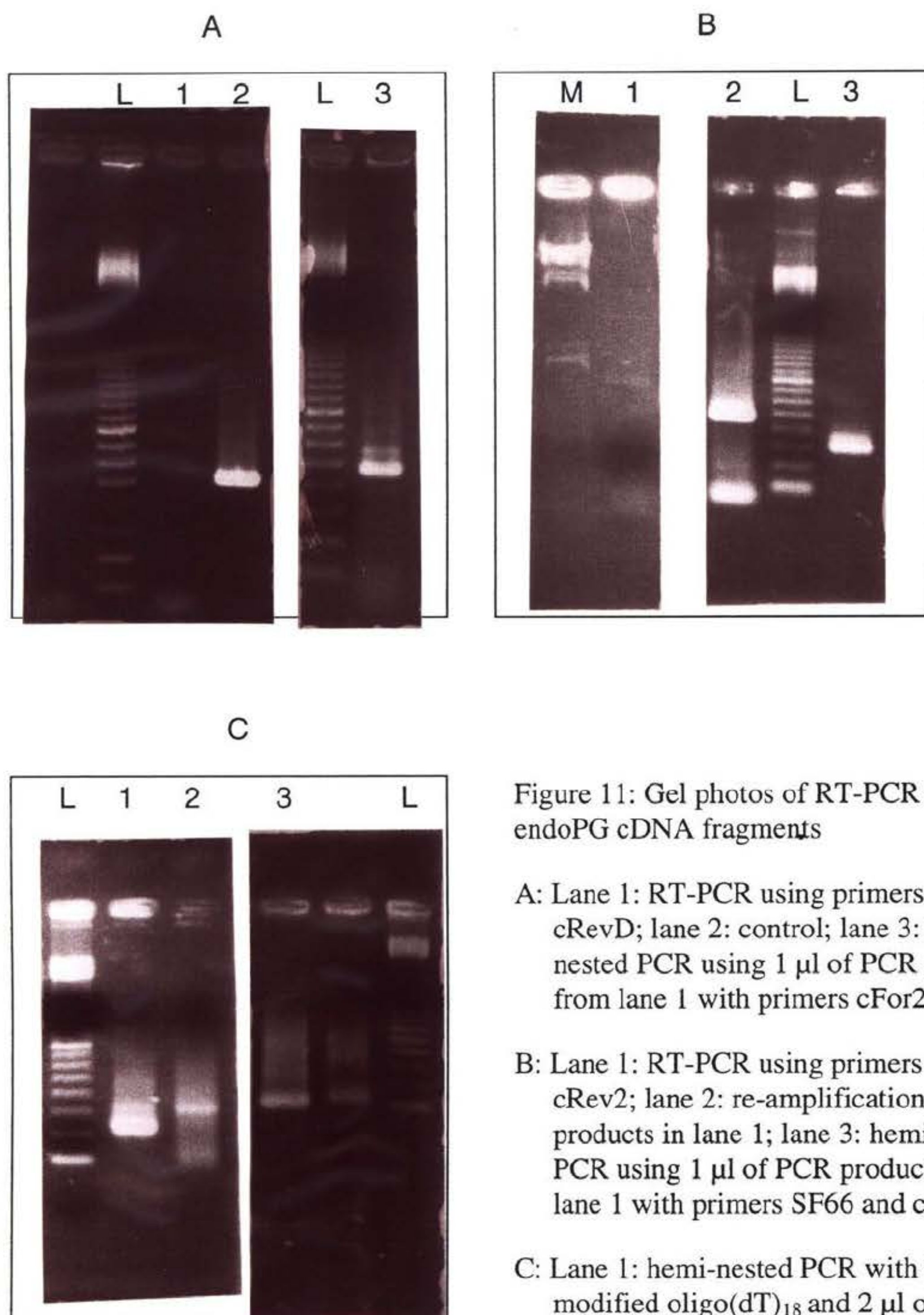


Figure 11: Gel photos of RT-PCR amplified endoPG cDNA fragments

A: Lane 1: RT-PCR using primers cFor2 and cRevD; lane 2: control; lane 3: hemi-nested PCR using 1 µl of PCR product from lane 1 with primers cFor2 and Con1.

B: Lane 1: RT-PCR using primers cForD and cRev2; lane 2: re-amplification of the products in lane 1; lane 3: hemi-nested PCR using 1 µl of PCR product from lane 1 with primers SF66 and cRev2.

C: Lane 1: hemi-nested PCR with SF10, modified oligo$(dT)_{18}$ and 2 µl of the product of RT-PCR, which was performed using primers cFor3 and modified oligo$(dT)_{18}$; lane 2: re-amplification of the products of RT-PCR mentioned above; lane 3: hemi-nested PCR with cFor3, cRev3 and 2 µl of the product of RT-PCR mentioned above.

L: 100 bp ladder, M: lambda DNA digested by Hind III

size (Fig.11, A2). To be sure that this was the authentic endoPG cDNA fragment, a heminested PCR was performed using the above PCR product as template and as primers ePGcFor2 with a primer upstream to the ePGcRevD, ePGcon1. A predominant DNA fragment of the expected size was produced together with several fragments of different sizes (Fig.11, A3). The predominant DNA fragment (about 500bp) from the heminested PCR was cloned and sequenced. This was the 5' half of the endoPG cDNA, termed cDNA1. Its sequence was colinear with the genomic gene sequence except that four introns were spliced out of it.

EndoPG cDNA1 together with the endoPG gene fragment amplified by RT-PCR using degenerate primers (these two cDNA fragments overlap) constituted a 552bp sequence starting from the putative translation initiation codon ATG of the endoPG cDNA. A total of 5 introns were spliced from this part of the cDNA.

2.2 RT-PCR amplification of the endoPG cDNA fragment 2

Three combination of primer pairs were used in RT-PCR amplification of the 3' half of the cDNA. All of them used primer cForD as the forward primer but different reverse primers: oligo$(dT)_{12-18}$, primer cRev1 and primer cRev2 from near the 3' end of the gene sequence. Different PCR conditions were used for the different primer pairs and heminested PCR was performed after the initial amplification. Only the reaction using the primer combinaton cForD with cRev2 and its subsequent heminested reaction amplified a predominant fragment of the predicted size (Fig. 11, B). Cycle sequencing with the purified DNA from the predominant band of the initial PCR comfirmed that it

was indeed the cDNA fragment. The cDNA sequence was colinear with the genomic gene sequence except that three introns were spliced out. This together with the cDNA sequence obtained previously constituted nearly the entire coding sequence of the endoPG gene. Eight introns were identified in the structural gene by this cDNA and genomic DNA sequence comparison.

This, however still left uncertain the identity of about 10 amino acids at the C-terminal of the encoded endoPG peptide. If oligo(dT) could be successfully used as the reverse primer, not only the end of the coding region of the gene but the actual polyadenylation site could be identified as well. Previous RT-PCRs using the oligo(dT)$_{12\text{-}18}$ primer as the reverse primer all failed. Hence a modified oligo(dT)$_{18}$ primer was designed with a stretch of 7 bp GC rich region added to the 5' end of the 18 bp T stretch in order to boost the specificity of the annealing of this reverse primer in the PCR amplification. This was used with two different forward primers in the RT-PCR amplification of endoPG cDNA fragment 3.

2.3 RT-PCR amplification of the endoPG cDNA fragment 3

Two primer combinations were used in the RT-PCR amplification of the endoPG cDNA fragment 3. These were: primer cForD with primer modified oligo(dT)$_{18}$ and primer cFor3 with modified oligo(dT)$_{18}$. After the original PCR a reamplification as well as the heminested PCR were performed for each. Gel analysis of the PCR products showed that

the primer combination cFor3 with the modified oligo$(dT)_{18}$ gave DNA fragments in the expected size range in both the reamplified original and its subsequent heminested PCR while nothing was amplified with the other primer pair. The products from the original as well as the two heminested PCR were more complex, revealing a smear with a predominant DNA band in the expected size range rather than just several distinctive DNA bands (Fig. 11, C). Due to the use of the nonspecific reverse primer modified oligo$(dT)_{18}$, the PCR products could have had much nonspecific amplifications making it difficult to purify the specific DNA band from the gel. After the reamplification of the original PCR products, the entire PCR mix was column purified to get rid of the primers and was used directly as template in cycle sequencing reaction. The internal primers (SF10, cRev3) were used. The sequence obtained was the same as that from the same region of the genomic sequence except that one intron was spliced from it, hence the last intron was uncovered. The endoPG cDNA fragment 3 sequence was thus obtained without cloning.

There was one thing peculiar about this sequencing reaction. Since a modified oligo$(dT)_{18}$ primer was used as the reverse primer in the amplification of the cDNA fragment 3, the sequence of the amplified cDNA fragment 3 was expected to end in polyA region but it was not so from this sequencing reaction. The 3' sequence was readable up to 80 bp past the stop codon. Then, the gel displayed bands in the position of all four bases for a stretch of about 30 nucleotides and the sequencing reaction stopped after that.

2.4 EndoPG cDNA sequence and its deduced peptide sequence

From the sequences of the three overlapping endoPG cDNA fragments, an endoPG cDNA sequence with an open reading frame of 1083 bps was obtained. From this, an endoPG peptide sequence was deduced (Fig. 12).

3.0 The relatedness of *C. purpureum* PG to PGs from other species

A dendrogram was constructed (Fig.13) from phylogenetic tree inference after aligning the deduced peptide sequence of this gene with the full-length peptide sequences of 16 PGs from plants, 15 PGs from fungi and 2 PGs from bacteria (Appendix 1). The dendrogram showed that the grouping of the enzyme was the same as the grouping of the taxa. The three major groups (fungi, plants and bacteria) were clearly distinguished by the relatedness of the enzymes. There are two types of PGs: endoPG and exoPG. Those definitely assigned as endoPG or exoPG were thus indicated after their species names in the dendrogram. Within plants or within fungi, endoPGs and exoPGs seemed to fall into two distinctive groups but because not every enzyme was assigned as endo- or exo-enzyme, it was not possible to conclude this yet. The Basidiomycete *Chondrostereum purpureum* PG was grouped with and was more related to the Ascomycete endoPGs and Ascomycete undefined PGs as compared to the two Ascomycete exoPGs who formed a separate group. This suggested strongly that the *C. purpureum* PG was an endoPG.

```
                                                                        18
5' ATG CCT TCG TTA TCC TCG ATC CTT AAG GGC CTT GGC GAC GTC ACC CTT TTT GCC
    M   P   S   L   S   S   I   L   K   G   L   G   D   V   T   L   F   A
                                                                        36
   TCT GTG GCT ACT GTA ACG GCC GTG CCG AAG CGA GCA TCC TGC ACC GTT GCC TCA
    S   V   A   T   V   T   A   V   P   K   R   A   S   C   T   V   A   S
                                                                        54
   GTT AGC GAT GCT GCA AAT ATT TCG GGA TGT ACC TCT GTG ACC ATA AAA TCG TTT
    V   S   D   A   A   N   I   S   G   C   T   S   V   T   I   K   S   F
                                                                        72
   ACC GTA CCT TCG GGC CAA ACT CTT GTT CTA AAC CCC TCA GAC GGT ACC ACC GTA
    T   V   P   S   G   Q   T   L   V   L   N   P   S   D   G   T   T   V
                                                                        90
   GCT ATG GTT GGC GAT GTC ACA TTT GCA AAG ACG ACA TCG TCT GGT CCT CTG TTT
    A   M   V   G   D   V   T   F   A   K   T   T   S   S   G   P   L   F
                                                                       108
   ACC ATT GAT GGC TCG AAC ATC AAA TTT AAA GGC GCA GGA CAT AAG TTC GAC GGA
    T   I   D   G   S   N   I   K   F   K   G   A   G   H   K   F   D   G
                                                                       126
   AAT GGC GCA AAG TAT TGG GAT GGG CAG GGG ACT AAC GGT GGC GTT ACC AAA TTT
    N   G   A   K   Y   W   D   G   Q   G   T   N   G   G   V   T   K   F
                                                                       144
   TAT CCT TTC CTC AAA TTC AAA GGA TCA GGT CAA TAC TCG AGT TTC ACT GTC CTA
    Y   P   F   L   K   F   K   G   S   G   Q   Y   S   S   F   T   V   L
                                                                       162
   AAC AGC CCT GCG CAG GCA ATA TCT ATC GGA AAT AGT GAC GGC CTA ACC TTT GAC
    N   S   P   A   Q   A   I   S   I   G   N   S   D   G   L   T   F   D
                                                                       180
   ACT GTC ACT GTG GAC AAT AGT GCC GGA GAT AGT GGT AGT CTC GGG CAC AAC ACT
    T   V   T   V   D   N   S   A   G   D   S   G   S   L   G   H   N   T
                                                                       198
   GAT GGC TTC GAT GTA TCT GCG GAC AAT GTT ACC ATC CAG AAT AGT GTT GTC AAA
    D   G   F   D   V   S   A   D   N   V   T   I   Q   N   S   V   V   K
                                                                       216
   AAC CAA GAT GAC TGC ATT GCC ATC AAC GAT CGG TCT AAT ATT GTT TTC CAA AAC
    N   Q   D   D   C   I   A   I   N   D   R   S   N   I   V   F   Q   N
                                                                       234
   AAT CAG TGC TCT GGA GGG CAC GGT ATT TCT GTT GGT TCT ATC GCC TCA GGA AAG
    N   Q   C   S   G   G   H   G   I   S   V   G   S   I   A   S   G   K
                                                                       252
   CAC GTA TCG GGT GTA GCA ATC AAA GGA AAC ACG GTC ACA AAC AGC ATG TAT GGC
    H   V   S   G   V   A   I   K   G   N   T   V   T   N   S   M   Y   G
                                                                       270
   ATG CGT GTC AAG GTC AAG GCT GCT GCA ACC TCT GCT TCA GTC TCT GCT GTG ACT
    M   R   V   K   V   K   A   A   A   T   S   A   S   V   S   A   V   T
                                                                       288
   TAC AGT GGC AAC ACC ATT TCT GGC ATT GCC AAG TAT GGC TTC CTC GTC TCG CAG
    Y   S   G   N   T   I   S   G   I   A   K   Y   G   F   L   V   S   Q
                                                                       306
   TCT TAC CCA GAT GAT GCT TCA ACT CCC GGA ACT GGC GCC CCA TCT CGG GGA TCA
    S   Y   P   D   D   A   S   T   P   G   T   G   A   P   S   R   G   S
                                                                       324
   ACT TCA GTG GTG ATA CTA CCA ATA TCA ATA AAC AGT GGA GCG AAA CGG GTC ACT
    T   S   V   V   I   L   P   I   S   I   N   S   G   A   K   R   V   T
                                                                       342
   GTC GAC TGT CGT AGT TGC ACC GGC ACG TGG AAC TGG TCT AAG CTG ACC GCT ACA
    V   D   C   R   S   C   T   G   T   W   N   W   S   K   L   T   A   T
                                                                       360
   GGG GGA AGT GCC GGC ACT ATC TCC AGT GAC AAA GCT AAG ATT TCT GGC GGC TCA
    G   G   S   A   G   T   I   S   S   D   K   A   K   I   S   G   G   S

   TAC TAG TTTAGTAGCCAAGCTCATGTTTGGTCAAGTTGTTACCCAGTTAGTGTTGTTGATTTAGCGTTG
    Y   *
   TAAATCCAGTCAATAATCTACCTGAATTGAATACTTTGTCAT 3'
```

Figure 12: cDNA sequence of the endoPG gene of *C. purpureum*. Putative N-glycosylation sites and polyadenylation site are underlined, polygalacturonase active site is double-underlined.

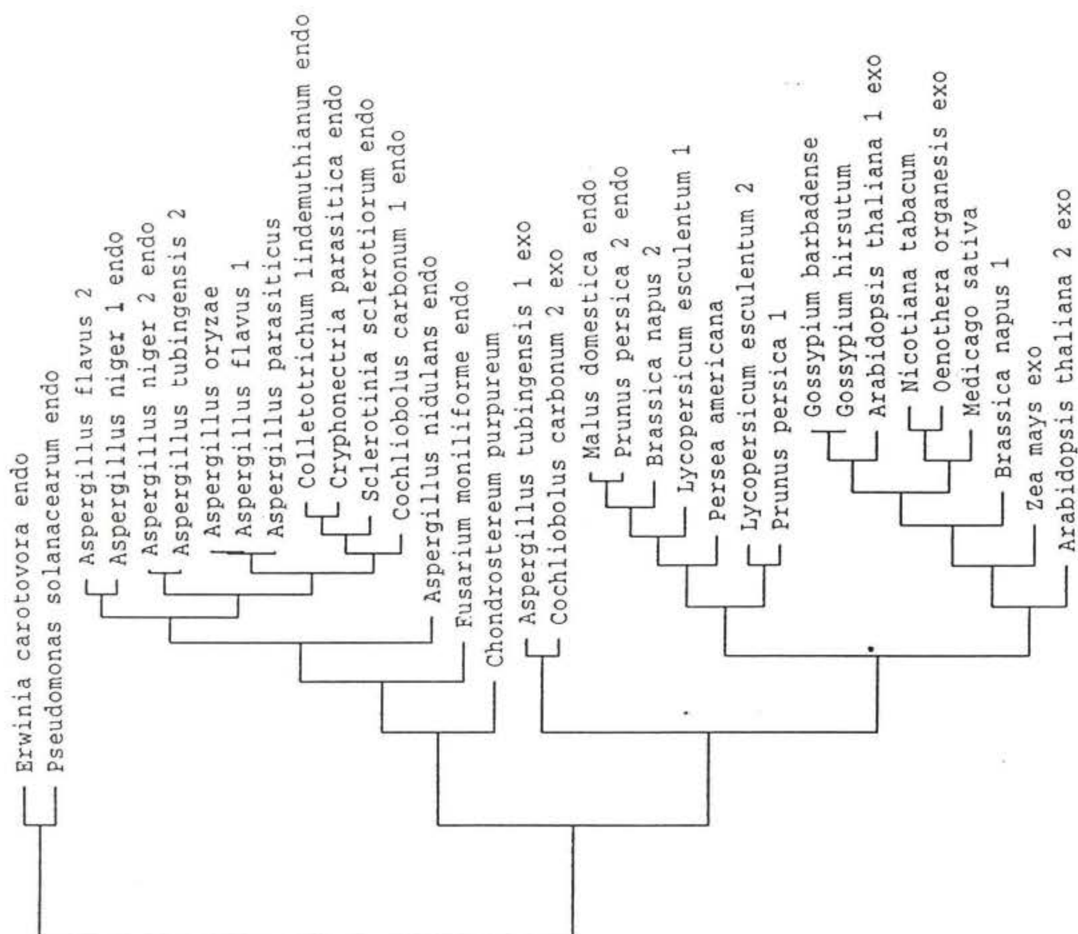


Figure 13: Dendrogram of relatedness of polygalacturonases from bacteria, fungi and plants

Discussion

Using primers designed from the conserved regions in the endoPG peptide alignment of bacteria, fungi and plants, a cDNA fragment of significant similarity at amino acid level with other fungal endoPGs was amplified from total RNA from *C. purpureum* by RT-PCR. A putative endoPG gene was isolated from the *C. purpureum* genomic library using this cDNA fragment as a probe. Both the genomic copy of the endoPG gene and the cDNA of the gene obtained by RT-PCR were sequenced and compared to other published sequences.

1.0 Characterization of the gene

1.1 Promoter and 5' untranslated region of the endoPG gene

The *Chondrostereum purpureum* endoPG gene has a coding region of 1596 bps with an open reading frame of 1083 bps which is interrupted by 9 introns (Fig. 14).

The assignment of the translation start site was based on the following: 1.) According to the comparison of the possible putative peptide sequences deduced from the endoPG gene with the consensus sequence from the endoPG peptide alignment, the possible peptide sequence encoded by the gene could be translated starting from position 1 or another ATG 409 bp upstream from this one. 2.) The forward primer designed from this ATG region used together with a reverse primer from the middle of the gene successfully amplified a cDNA fragment whereas the forward primer designed from the other ATG region (-409), failed to work.

```
                                GGGGGCTGGGNAGGCCCGTTGAAGGCATGGTT -1561
GGAGAGTGTGAAGGACATGGTTGAGTCCAAGTAGTGATATGCAATAATCAGACACTTGAT -1501
CCACGAATATATCACGTAACCCTAACCCGTCGTTAGCGCGTCAGACTTTCTGCTTGACCA -1441
CCATGTCAACTTCAGACTCTGAGGATGATGATGACTATAAACCATCCGAAAATGCTGGTG -1381
TGTTACTGAGTGTTTGGATACTGCTAGGCAATGCATTAACATGGTATAAGACGCTGAAAC -1321
GTCTGAAAATGAAGAACCGAGTGCAAAGCGGCCTAACGTTTCTACAGAGACTACTGAACT -1261
AGACCGAATACAGAAGAAACAGTAAGTTCGCCTTCCATTTGGTCTTGTTCATAGGTCATA -1201
ACTATCAGAATAGGGCACAAGATGCACTATGGGCCAACTTCACTGCATCAGTGAAGCAGC -1141
CTACAGCTAATATTTCTCCAGTGGCACAGAGAATGGTCAAAGTCCAGAAACGATATTTAT -1081
TCGCTGGTGGGGAAGTTACGTGCGTACATGAGGTTTCCGTGTGATACTTGGGGGACTAAT -1021
ATCCCGTCTGCCCAGAGAAGTGGTGGAGGTGGNGGAAAATTCAGATGATGCGAAGAAATG - 961
GCCTCTGGCGGTGCCCTCCTCCTCACTTGATGCAACTGAACCAGCTTCCAACACTGCAAA - 901
TATATGCCCTCCTGACCCAGTCAGCTCCTGCTCTAAAACCCCCACCACCGGAAAGGGCCT - 841
CGCAAGCCTAAGACCATCCTTTCTGCACTGCTTCGGACAAAGCCAAAAAAGCTTACAACA - 781
TTGGACAAATCTATGATGGACTGGAAAGCGCATCTCGACTCTACACCGGCAGAGGTCAAG - 721
GATGAATTAGATTCCAATCGCCGAGGGGGAGGGTATCTCGATAAAATCGATTTTCTACAG - 661
AGGGTGGAACATCGCAAGGAGAGTGTGCTTGAAGCCAACAAAGGAAAGAGAAGAAGAGGA - 601
TGAGTTGTGGATCCATATAGATATGTGAGCTCTCTCTGTTCTTTGACCGATACGACTAAA - 541
TACCGATAACCATAGTGTCAGTGTATTTGAGTCAGCCTAAATTGAATTTGAAATCGGCGT - 481
AAATTGACCCCCGGACGATCGTCAGTCAGTGTCCCCCCAACCAATGCTTATGATTTCGGG - 421
ATTCCGAATCACCTATGAATATCTATCTGGTCTTGTACTCTCGAACCCGACCTCTATACC - 361
AATTTCCTGTGATCCCCCGTTCGAATATTTCTAGTAGCCCTCGGCCAAGGAAGGACGATT - 301
GCCGCCCTTTGGTGGATGTTGCACGAGGTTGAGGCTTCCACATGCAGTAGGGATCCCCTA - 241
ACAGAGCTAATTTTATCCTTTTAGTGTCCGAAGCAGGAACAAGCAAGGCCCCCAGTACCT - 181
CGCGTTGACGACGGTCGATATTCGCGTCTTTTAAATTTGAGCTGCAAAGACCTGACCCAT - 121
GTAAAACGAGCTTTCCATATAAGAGGCTCTCATTCGAGGCATAAATCCGTTAATTTGTCT -  61
TCTCTCTCTCTCCCATCTCTCTCTCTTTCAAACACTCACTTGTTACATTCTTTCTTTACC -   1

ATGCCTTCGTTATCCTCGATCCTTAAGGGCCTTGGCGACGTCACCCTTTTTGCCTCTGTG     60
 M  P  S  L  S  S  I  L  K  G  L  G  D  V  T  L  F  A  S  V
                                                  .
GCTACTGTAACGGCCGTGCCGAAGCGAGCATCCTGCACCGTTGCCTCAGTTAGCGATGCT    120
 A  T  V  T  A  V  P  K  R  A  S  C  T  V  A  S  V  S  D  A

GCAAATATTTCGGGATGTACCTCTGTGACCATAAAATCGTTTACCGTACCTTCGGGCCgt    180
 A  N  I  S  G  C  T  S  V  T  I  K  S  F  T  V  P  S  G

gagtgctctcatttggttgtgcggttactgtggctcatcgaatccctagAAACTCTTGT    240
      Intron 1                                       Q  T  L  V

TCTAAACgtttgtacaccatgagaaaaccaatgctagtttctgattatctatccgacatg    300
 L  N                           Intron 2

ttagCCCTCAGACGGTACCACCGTAGCTATGGTTGGCGATGTCACATTTGCAAAGACGAC    360
    P  S  D  G  T  T  V  A  M  V  G  D  V  T  F  A  K  T  T

ATgtaggtggtctcttgttaaaaatgctatcccagttaacattctctattaagCGTCTGG    420
        Intron 3                                        S  S  G

TCCTCTGTTTACCATTGATGGCTCGAACATCAAATTTAAAGGCGCAGGACATAAGTTCGA    480
 P  L  F  T  I  D  G  S  N  I  K  F  K  G  A  G  H  K  F  D

CgtatgtttaagttcgcggaatcttcccttgtcttacctgacatttttgttcgtagGGA    540
                    Intron 4                                 G

AATGGCGCAAAGTATTGGGATGGGCAGGGGACTAACGGTGGCGTTACCAAATTTTATCCT    600
 N  G  A  K  Y  W  D  G  Q  G  T  N  G  G  V  T  K  P  H  P
```

```
TTCCTCAAATTCAAAGGATCAGGTCAATACTCGAGTTTCACTGTCCTAAACAGCCCTGCG   660
 F  L  K  F  K  G  S  G  Q  Y  S  S  F  T  V  L  N  S  P  A

CAGGCAATATCTATCGGAAATAGTGACGGCCTAACCTTTGACACTGTCACTGTGGACAAT   720
 Q  A  I  S  I  G  N  S  D  G  L  T  F  D  T  V  T  V  D  N

AGTGCCGGAGATAGTGGTAGTCTCGGGCACAACACTGATgtctgtacatttttctcagt   780
 S  A  G  D  S  G  S  L  G  H  N  T  D

gcgattacatctattcatgatacctacctttcctagGGCTTCGATGTATCTGCGGACAAT   840
  Intron 5                           G  F  D  V  S  A  D  N

GTTACCATCCAGAATAGTGTTGTCAAAAACCAAGATGACTGCATTGCCATCAACGATGGC   900
 V  T  I  Q  N  S  V  V  K  N  Q  D  D  C  I  A  I  N  D  G

TCTAATATTGTgtatgtgtgcctccttcttgtccgcaactcatccacgactggcttctta   960
 S  N  I  V                        Intron 6

ttgcccttgcagTTTCCAAAACAATCAGTGCTCTGGAGGGCACGGTATTTCTGTTGGTTC  1020
            F  Q  N  N  Q  C  S  G  G  H  G  I  S  V  G  S

TATCGCCTCAGGAAAGCACGTATCGGGTGTAGCAATCAAAGGAAACACGGTCACAAACAG  1080
  I  A  S  G  K  H  V  S  G  V  A  I  K  G  N  T  V  T  N  S

CATGTATGGCATGCGTGTCAAGGTCAAGGCTGCTGCAACCTCTGCTTCAGTCTCTGCTGT  1140
  M  Y  G  M  R  V  K  V  K  A  A  A  T  S  A  S  V  S  A  V

GACTTACAGTGGCAACACCATTTCTGGCATTGCCAAGTATGGCTTCCTCGTCTCGCAGTCT 1200
  T  Y  S  G  N  T  I  S  G  I  A  K  Y  G  F  L  V  S  Q  S

TACCCAGATGATGgtgcgtaacagagtcatattgaagagttattctgatttggcattata  1260
 Y  P  D  D                       Intron 7

tgggcagCTTCAACTCCCGGAACTGGCGCCCCATCTCGGGGATCAACTTCAGTGGTGATA  1320
     A  S  T  P  G  T  G  A  P  S  R  G  S  T  S  V  V  I

CTACCAATATCAAgtacgtcttccttcggtagtggatcgaatctcctttctcaatcttat  1380
 L  P  I  S                       Intron 8

tttttattattttacttagagTAAACAGTGGAGCGAAACGGGTCACTGTCGACTGTCGTA  1440
                      I  N  S  G  A  K  R  V  T  V  D  C  R

GTTGCACCGGCACGTGGAACTGGTCTAAGCTGACCGCTACAGGGGGAAGTGCCGGCACTA  1500
S  C  T  G  T  W  N  W  S  K  L  T  A  T  G  G  S  A  G  T

TCTCCAGTGACAAAGCTAAGgtactttgattcgaacgtactactgttagaaatgactgag  1560
I  S  S  D  K  A  K                Intron 9

cagccattctaacagATTTCTGGCGGCTCATACTAGTTTAGTAGCCAAGCTCATGTTTGG  1620
               I  S  G  G  S  Y

TCAAGTTGTTACCCAGTTAGTGTTGTTGATTTAGCGTTGTAAATCCAGTCAATAATCTAC  1680
CTGAATTGAATACTTTGTCATGAT                                      1704
```

Figure 14: Nucleic acid sequence of the endoPG gene of *Chondrostereum purpureum*
Pyrimidine-rich and GC-rich regions upstream of the coding sequence are underlined. putative TATA and CAAT boxes are double-underlined, a putative polyadenylation site is double-underlined.

3.) This ATG is embedded in the sequence TTTACC**ATG**C which is in agreement with the optimal context of an eukaryote initiation codon which consists of GCC(A,G) CC**AUG**G. The purine (A or G) 3 bases before the AUG and the G immediately after are considered to be the most important elements influencing the efficiency of translation by 10X (Lewin, B, 1994).

In the 5' non-transcribed region, several sequence motifs of possible functional significance were found: immediately upstream to the putative translation initiation codon, there is a region of about 70 bps which is composed predominantly of CT. Such pyrimidine rich regions in the promoter were also found in many other fungal genes. There are three putative TATA like sequence, TATAA, TAAAT and TAAT respectively within the distance of 103 bp upstream from the translation initiation codon ATG. Two CAAT like sequences, CAAA and CCAT lie within the correct distance upstream from the putative TATA sequences for them to serve as binding sites for regulatory proteins controlling transcription (Lewin, B., 1994). Further upstream, there is a GC rich region. Quite a few unusual sequences, like a stretch of one type of nucleotide, are also found in this promoter sequence.

1.2 The coding region of the gene

It is quite surprising that the coding region of this gene is interrupted by nine introns. In all the known fungal PG genes the intron numbers range from 0 to 4. The endoPG gene of *Sclerotinia sclerotiorum* has no introns (Reymond, P., et al., 1994) while the endoPG genes of *Cochliobolus carbonum* (Scott-Craig, J.S., et al., 1990) and of *Colletotrichum*

lindemuthianum (Centis, S., et al., 1996) contain one intron each. The (endo)PG genes I, II, III of *Aspergillus niger* contains two, one and three intron(s) respectively (Bussink, H.J.D., et al, 1992), and the endoPG gene from *Fusarium moliniforme* contains four introns (Caprari, C., et al., 1993). All of these fungi are ascomycetes. *Chondrostereum purpureum* is a Basidiomycete and hence the difference between Ascomycetes and Basidiomycetes may be reflected in the intron numbers contained in their genes coding the same protein.

The intron sequences in *C. purpureum* endoPG gene were compared to dertermine whether there are conserved sequences among the nine introns which could be important for the splicing. As expected, there was not much overall homology between these introns but certain sequences at the splice sites were conserved (Table 1). The putative lariate sequences among these nine introns were not well conserved and did not agree well with the consensus lariate sequence proposed for filamentous fungi (Gurr, S.J. et al, 1987). In higher eukaryotes, the lariate sequence is generally not conserved although statistically there is a strong preference for purine or pyrimidine in each nucleotide position of the lariate sequence. Only the nucleotide A at the lariate site is universally conserved (Lewin, B, 1994). This seemed to be the case with the *C. purpureum* endoPG gene.

Table 1: Sequence comparison of the nine introns of the *C. purpureum* endoPG gene

Intron No.	Intron splice sequence			Intron
	5'	Internal	3'	length (bp)
1	**GT**GA**GT**	GG**TTA**CT	T**AG**	52
2	**GT**TT**GT**	GA**TTA**TC	T**AG**	57
3	**GT**AG**GT**	AG**TTA**AC	A**AG**	51
4	**GT**AT**GT**	TC**TTA**CC	T**AG**	56
5	**GT**CT**GT**	GA**TTA**CA	T**AG**	57
6	**GT**AT**GT**	TC**TTA**TT	C**AG**	61
7	**GT**GC**GT**	AG**TTA**GT	C**AG**	64
8	**GT**AC**GT**	TC**TTA**TT	G**AG**	68
9	**GT**AC**TT**	TG**TTA**GA	C**AG**	54
Consensus*	GTANGT	YGCTAAC	YAG	< 100

*proposed for introns of filamentous fungi (Gurr at al. 1987)

The codon usage of this gene is summarized in Table 2.

1.3 The 3' untranslated region of the gene

In the 3' nontranslated region, there was a AATAAT sequence 74 bps downstream from the TAG stop codon which was similar to the polyadenynation signal AAUAAA found in the mRNA of higher eukaryotes.

2.0 One gene and at least four isoforms of the enzyme

In *Chondrostereum.purpureum*, it was reported that at least four isoforms of endoPG, glycoprotein having similar molecular weight but differing in their isoelectric points, exist in nature (Miyairi, et al., 1985). When a cDNA fragment of the putative endoPG gene was used as a probe to hybridize with the genomic DNA of *C. purpureum* digested by nine restriction enzymes individually, only one very strong hybridization signal was found in all the nine lanes (Fig. 9). This strongly suggests that there is a single copy of the endoPG gene in the *C. purpureum* genome. Furthermore only one copy of the cDNA fragment was amplified by RT-PCR using primers designed from the putative endoPG gene sequence. Taken together these data suggest that either there is only one copy of the endoPG encoding gene in *C. purpureum* or if there are more than one endoPG gene the other endoPG encoding genes are very different and have completely non-homologous sequences. In all the known cases in fungi where more than one endoPG encoding gene exist in the same genome, the similarity among the genes from the same genome are quite

Table 2: Codon usage of endoPG gene of *C. purpureum*

Amino acid	Codon	Counts	Norm. per 1000*	Norm. per family**	Amino acid	Codon	Counts	Norm. per 1000*	Norm. per family**
Gly	GGG	5.00	13.81	0.11	End	TGA	0.00	0.00	0.00
Gly	GGA	13.00	35.91	0.30					
Gly	GGT	8.00	22.10	0.18	Cys	TGT	2.00	5.52	0.33
Gly	GGC	18.00	49.72	0.41	Cys	TGC	4.00	11.05	0.67
Glu	GAG	0.00	0.00	0.00	End	TAG	1.00	2.76	1.00
Glu	GAA	0.00	0.00	0.00	End	TAA	0.00	0.00	0.00
Asp	GAT	11.00	30.39	0.52	Tyr	TAT	4.00	11.05	0.50
Asp	GAC	10.00	27.62	0.48	Tyr	TAC	4.00	11.05	0.50
Val	GTG	7.00	19.34	0.21	Leu	TTG	1.00	2.76	0.07
Val	GTA	6.00	16.57	0.18	Leu	TTA	1.00	2.76	0.07
Val	GTT	9.00	24.86	0.26	Phe	TTT	7.00	19.34	0.50
Val	GTC	12.00	33.15	0.35	Phe	TTC	7.00	19.34	0.50
Ala	GCG	3.00	8.29	0.10	Ser	TCG	10.00	27.62	0.19
Ala	GCA	9.00	24.86	0.30	Ser	TCA	10.00	27.62	0.19
Ala	GCT	10.00	27.62	0.33	Ser	TCT	16.00	44.20	0.30
Ala	GCC	8.00	22.10	0.27	Ser	TCC	3.00	8.29	0.06
Arg	AGG	0.00	0.00	0.00	Arg	CGG	3.00	8.29	0.50
Arg	AGA	0.00	0.00	0.00	Arg	CGA	1.00	2.76	0.17
Ser	AGT	11.00	30.39	0.21	Arg	CGT	2.00	5.52	0.33
Ser	AGC	3.00	8.29	0.06	Arg	CGC	0.00	0.00	0.00
Lys	AAG	11.00	30.39	0.52	Gln	CAG	5.00	13.81	0.56
Lys	AAA	10.00	27.62	0.48	Gln	CAA	4.00	11.05	0.44
Asn	AAT	8.00	22.10	0.38	His	CAT	1.00	2.76	0.25
Asn	AAC	13.00	35.91	0.62	His	CAC	3.00	8.29	0.75
Met	ATG	4.00	11.05	1.00	Leu	CTG	2.00	5.52	0.13
Ile	ATA	5.00	13.81	0.25	Leu	CTA	4.00	11.05	0.27
Ile	ATT	7.00	19.34	0.35	Leu	CTT	4.00	11.05	0.27
Ile	ATC	8.00	22.10	0.40	Leu	CTC	3.00	8.29	0.20
Thr	ACG	4.00	11.05	0.11	Pro	CCG	1.00	2.76	0.08
Thr	ACA	4.00	11.05	0.11	Pro	CCA	4.00	11.05	0.33
Thr	ACT	13.00	35.91	0.36	Pro	CCT	5.00	13.81	0.42
Thr	ACC	15.00	41.44	0.42	Pro	CCC	2.00	5.52	0.17
Trp	TGG	3.00	8.29	1.00					

*Codon frequency normalized per thousand codons

**Codon frequency normalized for each codon to a fraction within its synonymous family

high: at amino acid level, 60% sequence similarity are shared among three endoPG genes of *Aspergillus niger* (Bussink, et al, 1991b; Bussink et al, 1992a), 90% sequence identity between two (endo)PG genes of *Claviceps purpurea* (Tenberge, et al, 1996) and more than 90% sequence identity among three endoPG genes of *Sclerotinia sclerotiorum* (Reymond, et al, 1994). This supported the argument that only one endoPG encoding gene exists in the *C. purpureum* genome but definitive proof of this can only come from the knockout experiments of the endoPG gene. If the gene knockout abolishes the endoPG activity completely, it would mean that there is only one endoPG encoding gene in *C. purpureum*. This was a somewhat unusual result because of the four isoforms observed by Miyairi (1985). These isoforms of endoPG could arise from differential post-translational modification of one peptide, such as differential N-glycosylation, as is the case with the endoPG isozymes in *Fusarium moniliforme* (Caprari et al, 1993b). There are three potential N-glycosylation sites in the deduced endoPG sequence of *C. purpureum* (Fig. 12). Chemical or enzymatic deglycosylation of the endoPG isoenzyme mixture of *C. purpureum* can be used to test whether differential N-glycosylation is indeed responsible for the existence of the various isoforms.

3.0 Putative endoPG primary structure analysis, comparison with other PG sequences and secondary structure predictions

The deduced peptide sequence from the endoPG gene was analysed using computer program Gene Runner. A comparison of the amino acid composition of the deduced

Table 3: Comparison of the amino acid composition of the putative product of endoPG with the published EndoPG I of *C. purpureum*

Amino acid	endoPG product		published EndoPG I*	
	count	%	count	%
Ala	30	8.31	31	8.0
Arg	6	1.66	5	1.3
Asn	21	5.81	65	16.7
Asp	21	5.81		
Cys	6	1.66	6	1.5
Gln	9	2.49	14	3.6
Glu	0	0.00		
Gly	45	12.46	53	13.6
His	5	1.38	7	1.8
Ile	21	5.81	25	6.4
Leu	14	3.87	15	3.9
Lys	21	5.81	24	6.2
Met	4	1.10	3	0.8
Phe	13	3.60	12	3.1
Pro	12	3.32	22	5.7
Ser	53	14.68	28	7.2
Thr	36	9.97	43	11.1
Trp	3	0.83	4	1.0
Tyr	7	1.93	6	1.5
Val	34	9.41	26	6.7
Total	361	100.00	389	100.0

*Miyairi et al. 1985

endoPG with the amino acid analysis of purified endoPG I (Miraiyi, et al, 1985) is presented in Table 3. When the deduced peptide was compared with the endoPG1 characterized by Miyairi and colleagues (1985), certain discrepancies were found. The peptide deduced from this gene was a bit smaller than what was reported by Miyairi et al (1985), 361 a.a. versus 389 a.a.. Certain differences in the amino acid composition were also observed. The isoelectric point for the deduced protein was however similar to their result, which was pH 8.73 versus pH 8.5. As discussed earlier, there is an ATG site 409 bps upstream from the assumed translation start site which could serve as an alternate translation start site. If we take that ATG as the start site of the protein, the protein would be of 472 a.a. long which would be 83 a.a. longer than the result reported by Miyairi and colleagues (1985). Even if assuming that the cleavage of a secretion signal peptide occured to give rise to the mature protein, this ATG is unlikely to be the correct start signal as it would require an extraordinarily long peptide secretion signal. We concluded therefore that the deduced peptide is very likely the authentic gene product.

The consensus pattern of the polygalacturonase active site, [GSNH]-X(2)-[MFC]-X(2)-G-H-G-[LIVMAG]-X(1,2)-[LIVM]-G-S, where H is the active site residue, was proposed based on the best conserved region shared by PGs (endoPGs) and exoPGs from prokaryotic and eukaryotic origin. All sequences known to belong to this class and no other sequences were detected by the pattern (Generunner, 1994). The consensus pattern was found in this deduced peptide sequence, which is N-N-Q-C-S-G-G-H-G-I-S-V-G-S, confirming that the cloned gene is indeed a polygalacturonase encoding gene.

The full-length deduced endoPG peptide was used in BLAST search for similar protein sequences using all the protein data bases and the translated DNA data bases. The overall similarity among the PGs from bacteria, fungi and plants was quite low. Approximately 20% sequence similarity existed for pair-wise comparisons of the PG genes between bacteria and fungi, fungi and plants, or bacteria and plants. The deduced endoPG of *C. purpureum* has the highest sequence similarity with the endoPGs of *Sclerotinia sclerotiorum*, *Colletotrichum lindemuthianum*, *Cryphonectria parasitica* and *Cochliobolus carbonum* whereas its distance to the sequences of exoPGs of *Cochliobolus carbonum* and *Aspergillus tubingensis* was twice as large.

From the similarity comparison of the *C. purpureum* putative endoPG with the fungal endo- and exo- PGs, and from the relative distances of the PGs illustrated in Figure 13, it can be inferred that the putative endoPG of *C. purpureum* indeed belongs to the endo- rather than the exo- enzyme group. However, the definite assignment of the mode of action of this enzyme can come only from the actual enzymatic characterization.

Sequence alignment of PGs revealed that the region nearer to the C-terminal is better conserved as compared with the N-terminal region. There is a domain of about 90 amino acid which is conserved in all the 34 PGs (Fig. 15). The active site of the polygalacturonase is located within this conseved block of sequence. By chemical modification studies and an analysis of the kinetic parameters as a function of the pH (Cooke, et al, 1976; Rexova-Benkova and Mrackova, 1977; Miyairi, 1985), the essential amino acid for the action of polygalacturonase enzyme was suggested to be a histidine. Since there is only one universally conserved histidine among all the PGs, and it is located in the active

Organism	Sequence
Erwinia carotovora	NTDGIDPMSSKNITIAYSNIATGDDNVAIKAYKGRAETRNISILHNDFGT**GHGMSIGSET**-MGVYNVTVDDLKMNGTT-----NGLRIK
Pseudomonas solan.	NTDGFDPGQSTNVVLAYSYINTGDDHVAVKASSG--PTRNLLFAHNHFYY**GHGLSIGSET**NTGVSNMLVTDLTMDGNDSSAG-NGLRIK
Aspergillus tubing.	NTDGWDTYRSNNIVIQNSVINNGDDCVSFKPN-----STNILVQNLHCNG**SHGISVGSLG**QYKDEVDIVENVYVYNISMFNASDMARIK
Cochliobolus carb. 2	NTDGWDTYRSDNIIIQNSNINNGDDCVSFKPN-----STNILVQNLVCNG**SHGISVGSLG**QYPGEVDIVENILVRNISMSNASDGARIK
Chondrostereum purp.	NTDGFDV-SADNVTIQNSVVKNQDDCIAINDR-----S-NIVFQNNQCSG**GHGISVGSIA**---SG-KHVSGVAIKGNTVTNSMYGMRVK
Fusarium moniliforme	NTDGFDISSSDHVTLDNNHVYNQDDCVAVTSG-----T-NIVVSNMYCSG**GHGLSIGSVG**GK-SD-NVVDGVQFLSSQVVNSQNGCRIK
Aspergillus nidulans	NTDGFDIGESTYITITGAEIYNQDDCVAINSG-----E-NIYFSASVCSG**GHGLSIGSVG**GR-DD-NTVKNVTFYDVNVLKSQQAIRIK
Aspergillus niger 2	NTDAFDVGNSVGVNIIKPWVHNQDDCLAVNSG-----E-NIWFTGGTCIG**GHGLSIGSVG**DR-SN-NVVKNVTIEHSTVSNSENAVRIK
Aspergillus tubin. 2	NTDAFDVGNSVGVNIIKPWVHNQDDCLAINSG-----E-NIWFTSGTCIG**GHGLSIGSVG**GR-SN-NVVKNVTIEHSTVSNSENAVRIK
Aspergillus flavus 2	NTDAFDVSESNGVYITGANVKNQDDCLAINSG-----E-NIEFTGATCSG**GHGISIGSIG**NR-DT-NTVKNVKVADSTVVDSDNGIRIK
Aspergillus niger 1	NTDGFDISESTGVYISGATVKNQDDCIAINSG-----E-SISFTGGTCSG**GHGLSIGSVG**GR-DD-NTVKNVTISDSTVSNSANGVRIK
Sclerotinia scler.	NTDAFDVGSSTDITISGANVQNQDDCLAINSG-----T-GITFTGGTCSG**GHGLSIGSVG**GR-SD-NVVSDVIIESSTVKNSANGVRIK
Aspergillus flavus 1	NTDAFDVGSSTYINIDGATVYNQDDCLAINSG-----S-HITFTNGYCDG**GHGLSIGSVG**GR-SD-NTVEDVTISNSKVVNSQNGVRIK
Aspergillus oryzae	NTDAFDVGSSTYINIDGATVYNQDDCLAINSG-----S-HITFTNGYCDG**GHGLSIGSVG**GR-SD-NTVEDVTISNSKVVNSQNGVRIK
Aspergillus paras.	NTDAFDIGSSTYINIDGATVYNQDDCLAINSG-----S-HITFTNGYCDG**GHGLSIGSVG**GR-SD-NTVEDVTISNSKVVNSQNGVRIK
Cochliobolus carb. 1	NTDAFDIGSSSGITISNANIKNQDDCVAINSG-----S-DIHVTNCQCSG**GHGVSIGSVG**GR-KD-NTVKGVVVSGTTIANSDNGVRIK
Cryphonectria paras.	NTDAFDVGSSENIYISGAVINNQDDCLAINSG-----T-NITFTSGSCTG**GHGLSIGSVG**GR-SD-NTVKTVSITNSKIINSQNGVRIK
Colletotrichum lind.	NTDAFDVGSSTGVYISGADVKNQDDCLAVNSG-----T-NITFTGGTCSG**GHGLSIGSVG**GR-KD-NVVKSVSITNSKIINSDNGVRIK
Arabidopsis thal. 2	NTDGIHLSNADNVSILDSTIATGDDCVSVGRG-----SNNVTVERVICGP**GHGLSVGSLG**KYKNE-EDVSGIHVNNCTMIETDNGLRIK
Zea mays	NTDGIHMGDSSGITITNTVIGVGDDCISIGPG-----TSKVNITGVTCGP**GHGISIGSLG**RYKDE-KDVTDINVKDCTLKKTMFGVRIK
Brassica napus	NTDGIHLGRCEGVKILNTKIATGDDCISVGDG-----MKNLLIEKVVCGP**GHGISVGSLG**RYGWE-QDVTDITVKNCTLEGTSNGLRIK
Oenothera organesis	NTDGIHIGRSDGVNIINTEIKTGDDCISLGDG-----SKNINITNITCGP**GHGISVGSLG**RYKNE-ESVVGIYVKNCTITGSQNGVRIK
Nicotiana tabacum	NTDGIHVSRSSSVNITDSNFSTGDDCISVGDE-----TEQLYITRVTCGP**GHGISVGSLG**GNPDE-KPVVGVFVRNCTFTNTDNGVRIK
Gossypium barbadense	NTDGIHMGKSEGVNIIASDIKTGDDCISIGDG-----TKNMVIKEITCGP**GHGISIGSLG**KFQNE-EPVEGIKISNCTITNTSNGARIK
Gossypium hirsutum	NTDGIHMGKSEGVNIIASDIKTGDDCISIGDG-----TKNMVIKEITCGP**GHGISIGSLG**KFQNE-EPVEGIKISNCTITNTSNGARIK
Medicago sativa	NTDGIHMGKSTDVKILNTNIGTGDDCVSIGDG-----SKQITVQGVNCGP**GHGLSVGSLG**KFTTE-ENVEGITVKNCTLTATDNGVRIK
Arabidopsis thal. 1	NTDGIHIGRSNGVNLIGAKIKTGDDCVSIGDG-----TENLIVENVECGP**GHGISIGSLG**RYPNE-QPVKGVTVRKCLIKNTDNGVRIK
Lycopersicum esc. 1	NTDGIHVKSSSGVSIMKSQIGTGDDCISIGPG-----TSNLWIEGIACGP**GHGISIGSLG**WKQQE-LGVQNVTVKTVTFSGTTNGVRVK
Prunus persica 1	NTDGIHVQMSSGVTILNSKIATGDDCVSIGPG-----TSNLWIEGVACGP**GHGISIGSLG**KEQEE-AGVQNVTVKTVTFSGTQNGLRIK
Brassica napus 2	NTDGIHIVATKNIRISNSDIGTGDDCISIEDG-----SQNVQINDLTCGP**GHGISIGSLG**DDNSK-AYVSGINVDGATLSETDNGVRIK
Prunus persica 2	NTDGIHITNTKNITISSSVIGTGDDCISIVSG-----SQRVQATDITCGP**GHGISIGSLG**EDNAN-DHVSGVFVNGAKISGTSNGVRIK
Malus domestica	NTDGIHVTNTQNITISSSVIGTGDDCISIVSG-----SQRVQATDITCGP**GHGISIGSLG**EDGSE-DHVSGVFVNGAKLSGTSNGLRIK
Lycopersicum esc. 2	NTDGVHVSNTQYIQISDTIIGTGDDCISIVSG-----SQNVQATNITCGP**GHGISIGSLG**SGNSE-AYVSNVTVNEAKIIGAENGVRIK
Persea americana	NTDGIHITGTQRIHVMNSVIGTGDDCISIESG-----SKMVIATNITCGP**GHGISIGSLG**DRNSE-AHVSGVLVDGGNLFDTTNGLRIK

Figure 15: The best conserved amino acid domain in PG sequences

The proposed active site is in bold. The alignment was generated by MAP (Baylor's College of Medicine).

site of the enzyme, it is very likely that this histidine is involved in the catalytic reaction (Bussink, et al, 1991). This histidine is located at position 241 in the *C. purpureum* endoPG. Other than the active site, there is also a second motif, the positively charged residues Arg-Ile-Lys (RIK), that was considered to be specific to the PGs (Reymond et al, 1994). The RIK sequence, however, is not conserved in every PG whose sequence is known. RVK instead of RIK is found in one of the PGs produced by the plant *Lycopersicon esculentum*. In *C. purpureum*, RIK is also changed to RVK. This change from I (isoleusine) to V (valine) comes from the change in the first base of the codon- the change from G to A. The RIK motif was proposed (Bussink, et al, 1991) to be involved in the ionic interactions with the carboxylate groups present in the substrate since a chemical study of PG (Rexova-Benkova and Mrackova, 1977) indicated a carboxylate group as a possible component of the catalytic site.

The secondary structure of this endoPG was predicted using both Chou-Fasman and Garnier-Robson methods and very little α-helical structure was indicated by both methods. Garnier-Robson method predicted the protein to be composed entirely of β-sheet structure, as was reported for the secondary structure predictions of the other fungal endoPG sequences (Reymond P, et al., 1994). The putative secretion signal peptide of the endoPG is 25 amino acids long (Fig. 16).

Window	Subsequence	Probability
1 - 15	MPSLSSILKGLGD^VT	.000002
2 - 16	PSLSSILKGLGDV^TL	.000000
3 - 17	SLSSILKGLGDVT^LF	.000001
4 - 18	LSSILKGLGDVTL^FA	.000214
5 - 19	SSILKGLGDVTLF^AS	.000013
6 - 20	SILKGLGDVTLFA^SV	.000189
7 - 21	ILKGLGDVTLFAS^VA	.000028
8 - 22	LKGLGDVTLFASV^AT	.000004
9 - 23	KGLGDVTLFASVA^TV	.000371
10 - 24	GLGDVTLFASVAT^VT	.008371
11 - 25	LGDVTLFASVATV^TA	.000123
12 - 26	GDVTLFASVATVT^AV	.000139
13 - 27	**DVTLFASVATVTA^VP**	**1.000000 *****
14 - 28	VTLFASVATVTAV^PK	.000182
15 - 29	TLFASVATVTAVP^KR	.000682
16 - 30	LFASVATVTAVPK^RA	.000228
17 - 31	FASVATVTAVPKR^AS	.000000
18 - 32	ASVATVTAVPKRA^SC	.000022
19 - 33	SVATVTAVPKRAS^CT	.000244
20 - 34	VATVTAVPKRASC^TV	.001817
21 - 35	ATVTAVPKRASCT^VA	.000031
22 - 36	TVTAVPKRASCTV^AS	.000001
23 - 37	VTAVPKRASCTVA^SV	.000791
24 - 38	TAVPKRASCTVAS^VS	.000006
25 - 39	AVPKRASCTVASV^SD	.000014

Figure 16: Prediction of the the most likely cleavage site of a signal peptide.

For the computation of the probabilities of cleavage program SIGNAL was used. The program is part of HUSAR suite of GENIUS system in German Cancer Research Center (Heidelberg), it was written by Karl-Heinz Glatting.

4.0. Further considerations for the study of the role of endoPG in the pathogenicity of *C. purpureum*

As was already elaborated in the introduction of this thesis, the cloning of the endoPG gene is crucial for a refined assessment of the role of endoPG in the pathogenicity of *C. purpureum*. The endopolygalacturonase gene is the first protein encoding gene ever cloned from this fungus. With the cloning and characterization of the gene, the endoPG gene knockout and gene overexpressing experiments in *C. purpureum*, are now feasible. It is especially fortunate that there seems to be only one endoPG encoding gene in *C. purpureum* genome. Because *C. purpureum* also produce other pectic enzymes, it is important to be able to measure the enzyme activity specifically for endoPG in pathogenicity studies. Purified endoPG may therefore be indispensable. An easy method for column purification of specific protein relies on antibodies directed against the target protein. With the cloning and characterization of the endoPG gene we are in a position to design antibodies based on the primary structure of the deduced protein.

In order to engineer endoPG gene knockout strains and endoPG gene overexpressing strains, successful transformation of this fungus with a selectable marker gene is the next important step. Experiments directed towards transformation of *C. purpureum* with a dominant selectable marker are summarized in the Appendix 1.

Aside from pathogenicity studies, endoPG is an important enzyme in the food processing industry. If an endoPG overexpressing strain is obtained in *C. purpureum* and the fungus has aquired a GRAS (Generally Regarded As Safe) status, the endoPG overexpressing strain may be registered as an excellent source of the enzyme production. In one

industrially important filamentous fungus, *Aspergillus niger*, polygalacturonase-overproducing transformants have already been obtained by cotransformation (Bussink, et al, 1992b).

Bibliography

Agrios, G. N. 1978 Plant pathology, 2nd Edition. Academic Press, Inc.

Agrios, G. N. 1988 Plant pathology, 3rd Edition. Academic Press, Inc.

Archibold, O.W. 1979 Buried propagules as a factor in postfire regeneration in northern Saskatchewan. Canadian Journal of Botany 57:54-58.

Auld, B.A. 1990 Mycoherbicides: one alternative to chemical control of weeds. In: Alternative to chemical control of weeds: proceedings of an international conference, Rotorua, New Zealand. (Eds: Bassett, C., Whitehouse, L.J., and Zabkiewicz, J.A.). Ministry of Forestry, FRI Bulletin. PP. 71-73

Basham, H.G.and Bateman, D.F. 1975 Relationship of cell death in plant tissue treated with a homogeneous endopectate lyase to cell wall degradation. Physiological Plant Pathology 5:249-262.

Bateman, D.F. and Basham,H.G. 1976 Degradation of cell walls and membranes by microbial enzymes. In: Encycl Plant Physiol Vol. 4. (Ed by R. Heitefuss & P.H. Williams), pp. 316-355. Springer-Verlag, New York.

Benhamou, N., Laffite, C., Barthe, JP., Esquerre-Tugaye, MT. 1991 Cell surface interactions between bean leaf cells and *Colletotrichum lindemuthianum*. Cytochemical aspects of pectin breakdown and fungal endopolygalacturonase accumulation. Plant physiology 97: 234-244.

Bishop, G.C. 1979 Infection of cherry trees and production of a toxin that causes foliar silvering by different isolates of *Chondrostereum purpureum*. Australian Journal of Agricultural Research 30:659-665.

Braunholtz, J.T. 1981 Crop protection: the role of the chemical industry in an uncertain future. Philosophical Transactions of the Royal Society of London, Section B Biological Sciences 295:19-34.

Brooks, F.T. and Moore,W.C. 1926 Silverleaf disease V. Journal of Pomology 5:61-97.

Brown R.L., Cleveland T.E., Cotty P.J., Mellon J.E. 1992 Spread of *Aspergillus flavus* in cotton bolls, decay of intercarpellary membranes, and production of fungal pectinases. Phytopathology 82: 462-467.

Bussink, H.J.D., Kester, H.C.M.,Visser, J. 1990 Molecular cloning,nucleotide sequence and expression of the gene encoding prepropolygalacturonase II of *Aspergillus niger*. FEBS Letters 273:127-130.

Bussink, H.J.D., Buxton, F.P., Visser, J. 1991 Expression and sequence comparison of the *Aspergillus niger* and *Aspergillus tubigensis* genes encoding polygalacturonase II. Current Genetics 19:467-474.

Bussink, H.J.D., Brouwer, K.B., de Graaff, L.H., Kester, H.C.M., Visser, J. 1991b Identification and characterization of a second polygalacturonase gene of *Aspergillus niger*. Current Genetics 20:301-307.

Bussink, H.J.D., Buxton, F.P., Fraaye, B.A., de Graaff, L.H., Visser, J. 1992a The polygalacturonases in *Aspergillus niger* are encoded by a family of diverged genes. European Journal of Biochemistry 208 (1) : 83-90.

Bussink, H.J.D., van den Hombergh, J.P.T.W., van den IJssel, P.R.L.A., Visser, J. 1992b Characterization of polygalacturonase-overproducing *Aspergillus niger* transformants. Applied Microbiology and Biotechnology 37: 324-329.

Caprari, C., Richter, A., Bergmann, C., Cicero, S.L., Salvi, G., Cervon, F., Lorenzo, G.D.1993 Cloning and characterization of a gene encoding the endopolygalacturonase of *Fusarium moniliforme*. Mycological Research 97(4):497-505.

Caprari, C., Bergmann, C., Migheli, Q., Salvi, G., Albersheim, P., Darvill, A., Cervon, F., Lorenzo, G.D.1993b *Fusarium moniliforme* secretes four endopolygalacturonases derived from a single gene product. Physiological and Molecular Plant Pathology 43, 453-462

Centis, S., Dumas, B., Fournier, J., Marolda, M., Esquerre-Tugaye, MT. 1996 Isolation and sequence analysis of *Clpg1*, a gene coding for an endopolygalacturonase of the phytopathogenic fungus *Colletotrichum lindemuthianum*. Gene 170: 125-129.

Cervone, F., Hahn, M.G., Lorenzo,. G.D., Darvill, A., Albersheim, P. 1989 Host-pathogen interactions XXXIII. A plant protein converts a fungal pathogenesis factor into an elicitor of plant defense responses. Plant Physiology 90: 542-548.

Cervone, F., Lorenzo, G.D., Degra`, L., Salvi, G. 1987 Elicitation of necrosis in *Vigna unguiculata* walp. by homogeneous *Asperginus niger* endo-polygalacturonase and by α-D-galacturonate oligomers. Plant Physiology 85: 626-630.

Cervone, F., Lorenzo, G.D., Salvi, G., Camardella, L. 1986 Molecular evolution of fungal polygalacturonase. In: Biology and Molecular biology of Plant-Pathogen Interactions (ed. by J. Bailey) pp.385-392.NATO ASI Series H1 Springer-Verlag Berlin Heidelberg.

Cleveland, T.E., and Cotty P. J. 1991 Invasiveness of *Aspergillus flavus* isoates in wounded cotton bolls is associsted with production of a specific fungal polygalacturonase. Postharvest Pathology and Mycotoxins 81:155-158.

Chou P.Y. and Fasman, G.D. 1978 Prediction of protein secondary structure of proteins from their amino acid sequence. Advances in Enzymology 47: 45-147.

Cooke, R.D., Ferber, C.E.M., Kanagasabapathy, L. 1976 Purification and Characterisation of polygalacturonases from a commercial *Aspergillus niger* preparation. Biochimica et Biophysica Acta 452:440-451.

Cooper, R.M. 1977 Regulation of synthesis of cell-degrading enzymes of plant pathogens.In: cell wall biochemistry related to specificity in Host-plant pathogen Interactions (Ed. by B.Solheim & J. Raa), pp. 163-211. Universitetsforlaget, Oslo.

Cooper, R.M., Rankin, B., Wood, R.K.S. 1978 Cell Wall-degrding enzymes of vascular wilt fungi. II. Properties and modes of action of polysaccharidases of *Verticillium albo-atrum* and *Fusarium oxysporum f.sp. lycopersici*. Physiological Plant Pathology 13, 101-34.

Cooper, R.M. 1983 The mechanism and significance of enzymic degradation of host cell walls by parasites. In: Biochemical Plant Pathology (ed. by J. Callow), pp.101-135. John Wiley & Sons, New York.

Cooper, R.M. 1984 The role of cell wall-degrading enzymes in infection and damage.In: Plant diseases: infection, damage and loss(ed. by R.K.S. Wood & G.J.Jellis), pp.13-27. Blackwell Scientific Publications, Oxford.

Copas, L., and Williams, R.R. 1987 Effect of paclobutrazol on the inoculum potential of *Chondronstereum purpureum* and silverleaf symptoms in cider apple cvs Michelin and Sweet Coppin. Journal of Horticultural Society 62(3) 287-290.

Dayoff, M. O., Schwartz, R. M., and Orcutt, B. C. 1979 A model of evolutionary change in proteins. In Atlas of Protein Sequence and Structure, M. O. Dayoff, ed., National Biomedical Research Foundation, Washington D.C..

de Jong,M.D., Scheepens, P. C., Zadoks,J.C. 1990 Risk analysis for biological control: A Dutch case study in biocontrol of *Prunus serotina* by the fungus *Chondrostereum purpurem*. Plant Disease 74:189-194.

de Jong, M.D. 1992 Risk assessment for the application of biological control of a forest weed by a common plant pathogenic fungus. Society Risk Analysis 12: 465-466.

Favaron, F., Peretto, R., Bonfante, P., Alghisi P. 1993 Differental absorption and localization of two *Sclerotinia sclerotiorum* endo-polygalacturonases in soybean hypocotyls. Physiological and Molecular Plant Pathology 43: 353-364.

Feng, D-F and Doolittle, R. F. 1987 Progressive sequence alignment as a prerequisite to correct phylogenetic trees. Journal of Molecular Evolution 25:351-360.

Fielding, A.H.1981 Natural inhibitors of fungal polygalacturonases in infected fruits. Journal of General Microbiology 123:377-381.

Fincham, J.R.S.1989. Transformation in fungi. Microbiological Reviews 53:148-170.

Garnier, J., Osguthorpe, D.J. and Robson, B. 1978 Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. Journal of Molecular Biology 120: 97-120.

Gurr, S.J., Unkles, S.E., Kinghorn, J.R. 1987 The structure and organization of nuclear genes of filamentous fungi. Pages 93-139 in : Gene structure in eukaryotic microbes. J.R. Kinghorn, ed. IRL Press, Oxford, WA.

Hahn, M.G., Bucheli, P., Cervon,F., Doares,S.H., O'Neill,R.A., Darvill, A., Albersheim, P.1989 The role of cell wall constituents in plant-pathogen interactions. In T Kosuge, E Nester,eds, Plant-microbe Interactions, Vol 3. McGraw-Hill, New York.

Halleran, M. 1990 Forest vegetation management and the politics of the environment. Forestry Chronicle 66:369-371.

Hill, G.D. 1982 Impact of weed science and agricultural chemicals on farm productivity in the 1980's. Weed Science 30:426-429.

Kellman, M.C. 1970 Viable seed content of some forest soil in coastal British Columbia. Canadian Journal of Botany 48:1383-1385.

Kimura, M. 1983 The Neutral Theory of Molecular Evolution, Cambridge University Press, Cambridge.

Kusters-van Someren, M.A., Harmsen, J. A. M., Kester, H. C. M., Visser, J. 1991 Structure of the *Aspergillus niger* pelA gene and its expression in *Aspergillus niger* and *Aspergillus nidulans*. Current Genetics 20:293-299.

Kyte, J and Doolittle R.F. 1982 A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157:105-132.

Lewin, B. 1994 Genes V. Oxford University Press, Oxford, N.Y..

Maniatis, T., Frisch, E.F., Sambrook, J. 1982 Molecular cloning. A laboratory manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Mansfield, J.W. and Richardson, A.1981 The ultrastructure of interactions between *Botrytis* species and broad bean leaves. Physiological Plant Pathology 19:41-48.

McKeen,W.E. 1974 Mode of penetration of epidermal cell walls of *Vicia faba* by *Botrytis cinerea*. Phytopathology 64: 461-467.

Miyairi, K., Okuno, T., Sawai, K. 1985 Purification and properties of endopolygalacturonase I from *Stereum purpureum*, a factor inducing silver-leaf symptoms on apple trees. Agricultural and Biological Chemistry 49 (4) 1111-1118.

Miyairi, K., Okuno, T., Sawai, K.,1979 Studies on isoenzymes of the toxic endopolygalacturonase produced by *Stereum purpureum*. Bulletin of the Faculty of Agriculture of the Hirosaki University 31:1-10.

Mayairi, K. 1988 Biochemical studies on the silver-leaf inducing substance of apple silver-leaf disease. Bulletin of the Faculty of Agriculture of the Hirosaki University 49:145-148.

Mertz, L.M., Rashtchian, A. 1994 PCR radioactive labeling system: a rapid method for synthesis of high specific activity DNA probes. Focus 16 Number 2.

Mortensen, K. 1988 The potential of an endemic fungus, *Colletotrichum gloeosporioides*, for biological control of roundleaved mallow (Malve pusilla) and velvetleaf (Abutilon theophrasti). Weed Science, 36: 473-478

Needleman S. B and Wunsch C.D. 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48: 443-453.

Oswald, E.T. 1990 Theoretical overview of successional considerations in vegetation management. Forest Chronology 66:361-365.

Peace, T.R. 1962 Pathology of trees and shrubs. Clarendon Press,Oxford. 753 p.

Ramsfield, T. D., Becker, E. M., Rathlef, S. M., Tang, Y., Vrain, T. C., Shamoun, S. F., and Hintz, W. E. 1996 Geographic variation of *Chondrostereum purpureum* detected by polymorphisms in the ribosomal DNA. Canadian Journal of Botany 74:1919-1929

Rayner, A.D.M. 1977 Fungal colonization of hardwood stumps from natural sources. II. Basidiomycetes. Transaction of the British Mycological Society 69:303-312.

Rexova-Benkova, L., Mrackova, M. 1978 Active groups of extracellular endo-D-Galacturonase of *Aspergillus niger* derived from PH effect on kinetic data. Biochimica et Biophysica Acta 523:162-169.

Reymond, P., Deleage, G., Rascle C., Fevre, M. 1994 Cloning and sequence analysis of a polygalacturonase-encoding gene from the phytopathogenic fungus *Sclerotinia sclerotiorum*. Gene 146:233-237.

Ridings, W.H. 1986 Biological control of stangler vine in citrus-a researcher's view.Weed Science 34(Suppl. 1):31-33.

Ruttkowski, E., Labitzke, R., Khanh, N.Q., Loffler, F., Gottschalk, M. Jany, K-D. 1990 Cloning and DNA sequence analysis of a polygalacturonase cDNA form *Aspergillus niger* RH5344. Biochi et Biophysic Acta 1087:104-106.

Saitou, N. and Nei, M. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Schafer, W. 1994 Molecular mechanism of fungal pathogenicity to plants. Annual Review of Phytopathology 32: 461-77

Scheepens, P.C., and Hoogerbrugge, A. 1988 Bestrijding van Amerikaanse vogelkers met loodglansschimmel(*Chondrostereum purpureum*). Gewasbescherming 19:141-147.

Scheepens, P.C. 1980 Bestrijding van de Amerikaanse vogelkers met pathogen schimmels; een perspectief. Cent Agrobiol Onderz, Wageningen. Verslag Nr. 29.

Schuren, F.H.J., Wessels, J.G.H. 1994 Highly-efficient transformation of the homobasidiomycete *Schizophyllum commune* to phleomycin resistance. Current Genetics 26: 179-183.

Scott-Craig, J.S., Panaccione, D.G., Cervone, F., Walton, J.D. 1990 Endopolygalacturonase is not required for pathogenicity of *Cochliobolus carbonum* on maize. The Plant Cell 2: 1191-1200.

Setliff, E.C. and Wade E.K. 1973 *Stereum purpureum* associated with sudden decline and death of apple trees in Wisconsin. Plant Disease Report 57:473-474.

Shaw,W.C. 1978 Herbicides: the cost benefit ratio-the public view. Proceedings of the South Weed Science Society 31:28-47.

Sneath, P.H.A. and Sokal, R.R. (1973) in Numerical Taxonomy (pp; 230-234), W.H. Freeman and Company, San Francisco, California, USA)

Spiers, A.G., and Hopcroft, D.H. 1988 Ultrastructural studies on basidial and basidiospore development and basidiospore release in *Chondrostereum purpureum*. European Journal for Pathology 18:367-381.

Studier J.A. and Keppler K. J. 1988 A note on the neighbor-joining algorithm of Saitou and Nei. Molecular biology and evolution 5:729-731.

Talmadge, K.W., Keegstra, K., Bauer, W.D., Albersheim, P. 1973 The structure of cell walls. I. The macromolecular components of the walls of suspension-cultured sycamore cells with a detailed analysis of the pectic polysaccharides. Plant Physiology 51: 158-173.

Templeton,G.E., Smith R.J., Klomparens, W. 1980 Commercialization of fungi and bacteria for biological control. Biocontrol News and Information 1:291-294.

Templeton,G.E., TeBeest, D.O., Smith, R.J. 1984 Biological weed control in rice with a strain of *Colletotrichum gloeosporioides* (Penz.) Sacc. used as a mycoherbicide. Crop Protection 3:409-422.

Templeton, G.E. 1986 Mycoherbicide research at the university of Arkansas-past, present, and future.Weed Science 34(Suppl. 1):35-37.

Tenberge, K.B., Homann, V., Oeser, B., Tudzynski, P. 1996 Structure and expression of two polygalacturonase gene of *Claviceps purpurea* oriented in Tandem and cytological evidence for pectinolytic enzyme activity during infection of rye. Phytopathology Vol.86, No.10:1084-1097.

Timberlake,W.E., Boylan,M.I., Cooley, M.B., Mirabito, P.M., O'Hara, E.B.,Willet C.E. 1985 Rapid identification of mutation-complementing restriction fragments from *Aspergillus nidulans* cosmids. Experimental Mycology 9:351-355.

Wahlstrom, K., Karlsson, J-O., Holdenrieder, O., and Stenlid, J. 1991. Pectinolytic activity and isozymes in European *Armillaria* species. Canadian Journal of Botany, 69: 2732-2739.

Wall R.E., Prasad R., Shamoun, S.F. 1992 The development and potential role of mycoherbicides for forestry. Forestry Chronicle 68: 736-741.

Wall, R.E.1990 The fungus *Chondrostereum purpureum* as a silvicide to control stump sprouting in hardwoods. North Journal of Applied Forestry 7:17-19.

Wall, R.E. 1986 Pathogenicity of *Chondrostereum purpureum* to yellow birch. Plant Disease 70:158-160.

Ward, M. 1991 *Aspergillus nidulans* and other filamentous fungi as genetic systems. In: Modern Microbial Genetics pp455-496. Wiley-Liss, Inc.

Watson, A.K. 1989 Current advances in bioherbicide research.pp.987-996. Proceedings of the Brighton Crop Protection Conference - Weeds.

Wernars, K., Goosen, T., Wennekes, B.M.J., Swart, K., van den Hondel, C.A.M.J.J.,van den Broek, H.W.J.1987 Cotransformation of *Aspergillus nidulans*: A tool for replacing fungal genes. Molecular and General Genetics 209:71-77.

Wilson, B.F. 1968 Red maple stump sprouts: development the first year. Harvard Forest Papers No. 18.

Woodgate-Jones, Phillipa and Hunter, T. 1983 Integrated control of *C. purpureum* in plum by treatment of pruning wounds. Journal of Horticural Science 8:491-495.

Appendix 1

Result of Multiple Peptide Sequence Alignment of the following 34 Polygalacturonases Using PILEUP

erw-caro: *Erwinia carotovora* endo

pse-sola: *Pseudomonas solanacearum* end

asp-fla2: *Aspergillus flavus* 2

asp-nig1: *Aspergillus niger* 1 endo

asp-nig2: *Aspergillus niger* 2 endo

asp-tub2: *Aspergillus tubingensis* 2

asp-ory: *Aspergillus oryzea*

asp-fla1: *Aspergillus flavus* 1

asp-par: *Aspergillus parasiticus*

col-lin: *Colletotrichum lindemuthianum* endo

crp-par: *Cryphonectria parasitica* endo

coc-car1: *Cochliobolus carbonum* 1 endo

scl-scle: *Sclerotinia sclerotiorum* endo

asp-nidu: *Aspergillus nidulans* endo

fus-mon: *Fusarium moniliforme* endo

chon-pur: *Chondrostereum purpureum*

asp-tub1: *Aspergillus tubingensis* 1 exo

coc-car2: *Cochliobolus carbonum* 2 exo

mal-dome: *Malus domestica* endo

pru-per2: *Prunus persica* 2 endo

bra-nap2: *Brassica napus* 2

lyc-esc1: *Lycopersicum esculentum* 1

per-amer: *Persea americana*

lyc-esc2: *Lycopersicum esculentum* 2

pru-per1: *Prunus persica* 1

gos-bar: *Gossypium barbadense*

gos-hir: *Gossypium hirsutum*

nic-taba: *Nicotiana tabacum*

oen-org: *Oenothera organesis* exo

med-sat: *Medicago sativa*

bra-nap1: *Brasica napus* 1

ara-tha1: *Arabidopsis thaliana* 1 exo

zea-mays: *Zea mays* exo

ara-tha2: *Arabidopsis thaliana* 2 exo

```
          1                                                 50
erw-caro  .......... .......... .......... .......... ..........
pse-sola  .......... .......... .......... .......... ..........
asp-fla2  .......... .......... .......... .......... ..........
asp-nig1  .......... .......... .......... .......... ..........
asp-nig2  .......... .......... .......... .......... ..........
asp-tub2  .......... .......... .......... .......... ..........
 asp-ory  .......... .......... .......... .......... ..........
asp_fla1  .......... .......... .......... .......... ..........
 asp-par  .......... .......... .......... .......... ..........
 col-lin  .......... .......... .......... .......... ..........
 crp-par  .......... .......... .......... .......... ..........
coc-car1  .......... .......... .......... .......... ..........
scl-scle  .......... .......... .......... .......... ..........
asp-nidu  .......... .......... .......... .......... ..........
 fus-mon  .......... .......... .......... .......... ..........
chon-pur  .......... .......... .......... .......... ..........
asp-tub1  .......... .......... .......... .......... ..........
coc-car2  .......... .......... .......... .......... ..........
mal-dome  .........M ALKTQLLWSF VVVFVVSFST TSCSGSSFQE VNALHSYVDH
pru-per2  .........M ALQKHLVLFY V...VVSFCA ASCYSSGFQE VNSLHSFVDH
bra-nap2  .........M ARCHGSLAIF LCVLLM.... .......LAC CQALSSNVDD
lyc-esc1  .........M VIQRNSILLL IIIFASSIST ....CRSNVI DDNLFKQVY.
per-amer  .......... .MALTRLLLP ISILWFCFYS ....SHTILQ KDPLLICVN.
lyc-esc2  .......... .....MSPFA IFFLLLIYS. .......... ..........
pru-per1  .........M ANRRSLFSLS LIFVFMINS. .......... ..........
 gos-bar  .........M APHLNIVPSM FVLLLLFIS. .......... ..........
 gos-hir  .........M APHLNIVPSM FVLLLLFIS. .......... ..........
nic-taba  .......... .MDLKFKVHF ALVLLFLAH. .......... ..........
 oen-org  .......... .......... .......... .......... ..........
 med-sat  .......... ...MKFSTAI IVSFLFIAD. .......... ..........
bra-nap1  .........M GSYLGIYTIL VLCLLGYSA. .......... ..........
ara-tha1  .........M VGSHKASGVL LVLLVVMATT IANGTPVVDK AKNAATAVED
zea-mayx  .........M ACTNNAMRAL FLLVLFCI.. .......... ..........
ara-tha2  MANARSLVAK ANNINVGSLI LMALVF.... .......... ..........
```

```
          51                                                  100
erw-caro  .......... .......... .........M EYQSGKRVLS LSLGLIGLFS
pse-sola  .......... ...MNHRYTL LALAAAALSA GAHATGTSVT APWGEVAEPS
asp-fla2  .......... .......... .......... MH.FQLLVLA PLALSSAAAP
asp-nig1  .......... .......... .......... MHSYQLLGLA AVGSLVSAAP
asp-nig2  .......... .......... .........M HSFASLLAYG LVAGATFASA
asp-tub2  .......... .......... .........M HSFASLLAYG LAASATLASA
 asp-ory  .......... .......... ........MQ LLQSSVIAAT VGAALVAAVP
asp_fla1  .......... .......... ........MQ LLQSSVIAAT VGAALVAAVP
 asp-par  .......... .......... ........MQ LLQSSVIAAT VGAALVAAAP
 col-lin  .......... .......... ........MV SYLFVLGALA SVAIASPVPE
 crp-par  .......... .......... ...MFSTLLL AALLPLIQAA PAPAVTPAAH
coc-car1  .......... .......... .........M VAYALTSMLL SAGALVAAAP
scl-scle  .......... .........M VEILSSALSL LRLGAAVSAA PAPAPTAAPN
asp-nidu  .......... .......... ....MVRQLI LISSLLAAVA VRAPADPAHP
 fus-mon  .......... .......... ...MVRNIVS RLCSQLFALP SSS.......
chon-pur  .......... .......... ........MP SLSSILKGLG DVTLFASVAT
asp-tub1  ...MRLTHVL SHTL..GLLA LGATAEAFSR SREAACGPKK PFRPLP....
coc-car2  ...MRVTDII SCALLQASIA LSTPVEELG. ...AKAVVAK RFPPVPFLPG
mal-dome  VDDKESG.YN SRAYPSYTDT IEGLKVMELI RPRTQLFSSR KLNTITGGIA
pru-per2  ..EKESG.YN SRAHPSNMNT IEGVKFMEFI KPRAQLFSSR KLERAGS...
bra-nap2  GYGHEDGSFE TDSLIKLNND DDVLTLKSSD RPTTE..... ..........
lyc-esc1  ....DN..IL EQEFAHDFQA Y....LSYLS K......... NIESNNNIDK
per-amer  ....GDPGFD QRAYPTYFGP I....LDEFS SIMGFEPSIL SLERFNPVGG
lyc-esc2  .......... .......... .......... .......... ..........
pru-per1  .......... .......... .......... .......... ..........
 gos-bar  .......... .......... .......... .......... .........A
 gos-hir  .......... .......... .......... .......... .........A
nic-taba  .......... .......... .......... .......... .........F
 oen-org  .......... .......... .......... .......... ..........
 med-sat  .......... .......... .......... .......... .........F
bra-nap1  .......... .......... .......... .......... .........N
ara-tha1  TAKNAATAVG GAA....... .......... .......... ...ASVGAKV
zea-mayx  .......VHG EKE....... .......... .......... ...ESKGIDA
ara-tha2  ...GSCVANG EYL....... .......... .......... ...GGRRGLA
```

```
          101                                                  150
erw-caro  ASAWASDSRT VSEPKTP... ..SSCTTLKA DSSTATSTIQ KALNNCDQGK
pse-sola  LPADSAVCKT LSASITPIKG SVDSVDGNPA NSQPDASRIQ SAIDNCPAGQ
asp-fla2  A...PSRTSE LVERGSSCTF ...TSAAQAS ASAKSCSNIV LKNIAVPAGE
asp-nig1  A...PSRVSE FAKKASTCTF ...TSASEAS ESISSCSDVV LSSIEVPAGE
asp-nig2  S...P..... .IEARDSCTF ...TTAAAAK AGKAKCSTIT LNNIEVPAGT
asp-tub2  S...P..... .IEARGSCTF ...KTAAAAK AGKAGCSTIT LDNIEVPAGT
 asp-ory  V...E..... .LKARDSCTF ...TSAADAK SGKTSCSTIT LSNIEVPAGE
asp_fla1  V...E..... .LEARDSCTF ...TSAADAK SGKTSCSTIT LSNIEVPAGE
 asp-par  V...E..... .LEARDSCTF ...TSAADAK SGKTSCSTIT LSNIEVPAGE
 col-lin  L...K..... .ARA..SCTF ...TDAASAI KGKASCTTIV LNNIAVPAGT
 crp-par  L...E..... .DRASKSCTF ...TDAAAVS KSKASCATIT LNNIAVPSGT
coc-car1  S...G..... .LDARDGCTF ...TDAATAI KNKASCSNIV ISGMTVPAGT
scl-scle  V...ADALAA VEKRAGSCTF SGSSGAAAAI KSKASCATIV ISAVAVPSGT
asp-nidu  MVTEAPDVNL VEKRATTCTF SGSEGASKAS KSKTSCSTIY LSDVAVPSGT
 fus-mon  .......... .LQERDPCSV TEYSGLATAV S...SCKNIV LNGFQVPTGK
chon-pur  VTA....... .VPKRASCTV ASVSDAANIS ....GCTSVT IKSFTVPSGQ
asp-tub1  ....TSQSRD KTCHVRSHGD GTDDSDYILS ALNQ...CNH GGKVVFDEDK
coc-car2  KASSVPGSRN KTCMLKALGG GKDDSANILS AVKQ...CNN GGHVVFPKGQ
mal-dome  TSSAPAKTIS VDDFGAKGNG .ADDTQAFVK AWKAACSSSG AMVLVVPQ.K
pru-per2  KSSSSVKTIS VANFGAKGNG .ADDTRAFEK AWKAACSSNG AIVLVVPQ.K
bra-nap2  .....SSTVS VSNFGAKGDG KTDDTQAFKK AWKKACSTNG VTTFLIPKGK
lyc-esc1  VDKNGIKVIN VLSFGAKGDG KTYDNIAFEQ AWNEACSSRT PVQFVVPKNK
per-amer  PETSPDTDIS VDDFGARGDG .TDDTKAFEK AWKDACSSGS V..LIVPENK
lyc-esc2  SLAANTNIYN VQNYGAKSDG KTDSSKAFLN AWAAACASNK PSTINVPIG.
pru-per1  AIAS.PLTYN VASLGAKADG KTDSTKAFLS AWAKACASMN PGVIYVPAG.
 gos-bar  SKVQPDAFDV VAKFGAKADG KTDLSKPFLD AWKEACASVT PSTVVIPKG.
 gos-hir  SKVQSDAFDV VAKFGAKADG KTDLSKPFLD AWKEACASVT PSTVVIPKG.
nic-taba  GESQTGVFDI TK...YGANS NADISEALLN AFKEACQSTS PSTIVIPKG.
 oen-org  .......... .......... ..DSTQALTT AWKEACASAS PSTILVPKG.
 med-sat  CAAQSGVLDI SK...FGGKP NSDIGQALTS AWNEACASTT AAKIVIPAG.
bra-nap1  AEV....... ...FTAGGPP NSDITAAVLK AFTSACQAPA PSQVLIPKG.
ara-tha1  SGAKPGAAVD VKASGAKGDG KTDDSAAFAA AWKEACA..A GSTITVPKG.
zea-mayx  KASGPGGSFD ITKLGASGNG KTDSTKAVQE AWASACGGTG KQTILIPKG.
ara-tha2  ANSGNPTVYD ITKFGAVGDG STNTFKAFLN TWIQVCDSPV PATLLVPKG.
```

```
            151                                                    200
erw-caro    AVRLSAGST. .SVFLSGPLS LPSGVSLLID KGVTLRAVNN AKSFENAPSS
pse-sola    AVKLVKGSAG ESGFLSGSLK LKSGVTLWID TGVTLFASRN PADYDNGLGT
asp-fla2    TLDLSKAKDG ATITFEGTTT FGYK.EWKGP .......... ........LI
asp-nig1    TLDLSDAADG STITFEGTTS FGYK.EWKGP .......... ........LI
asp-nig2    TLDLTGLTSG TKVIFEGTTT FQYE.EWAGP .......... ........LI
asp-tub2    TLDLTGLTSG TKVIFEGTTT FDYE.EWAGP .......... ........LI
asp-ory     TLDLTGLNDG TTVIFSGETT FGYK.EWEGP .......... ........LI
asp_fla1    TLDLTGLNDG TTVIFSGETT FGYK.EWEGP .......... ........LI
asp-par     TLDLTGLNDG TTVIFSGETT FGYK.EWEGP .......... ........LI
col-lin     TLDMTGLKSG THVSFSGKTT FGYK.EWEGP .......... ........LI
crp-par     TLDLTKLNSG TKVIFAGTTS FGYK.EWEGP .......... ........LI
coc-car1    TLDLTGLKSG ATVTFQGTTT FGYK.EWEGP .......... ........LI
scl-scle    TLDLTGLKSG THVVFEGTTT FGYE.EWYGT .......... ........LV
asp-nidu    TLDLSDLNDG THVIFQGETT FGYE.EWEGP .......... ........LV
fus-mon     QLDLSSLQND STVTFKGTTT FATTADNDFN .......... ........PI
chon-pur    TLVLNP.SDG TTVAMVGDVT FA.KTTSSGP .......... ........LF
asp-tub1    EYIIGTALNM TFL...KNID LEVLGTILFT NDTDYWQANS FKQGFQNATT
coc-car2    QFTIGTALDL TFL...NGID LDIQGTIQFT NDTDYWQANS FKQVFQNATT
mal-dome    NYLVRPIEFS G.PCKS.QLT LQIYGTIEAS EDRSIY..KD IDHWL....I
pru-per2    TYLVRPIEFS G.PCKS.HLT MQIYGTIEAS DDRSVY..KD VTHWL....I
bra-nap2    TYLLKSIRFR G.PCKS.LRS FQILGTLSAS TKRSDY.SND KNHWL....I
lyc-esc1    NYLLKQITFS G.PCRS.SIS VKIFGSLEAS SKISDY..KD RRLWI....A
per-amer    NYLLKQITFS G.PCKS.DLR VKIRGTIEAS SDQSDWVGHN RKRWI....E
lyc-esc2    KYLIHNANFN GQTCKSKAIT MHIDGTLLAP SDYNVI..GN EENWI....K
pru-per1    TFFLRDVVFS G.PCKNNAIT FRIAGTLVAP SDYRVI..GN AANWI....F
gos-bar     TYLLSKVNLE G.PCKA.PIE INVQGTIQAP ADPSAF..KD PN.WV....R
gos-hir     TYLLSKVNLE G.PCKA.PIE INVQGTIQAP ADPSAF..KD PN.WV....R
nic-taba    TFTMNQVKLE G.PCKS.PLE LQIQATLKAP SDPSQL..KV GE.WL....T
oen-org     NFAVGLITLE G.PCKS.SIG LQLQGTLKAP ADPSKI..KG LG.WI....N
med-sat     TYQLNGIELK G.PCKA.PIE LQVDGTIQAP ADPSVI..KG TEQWF....K
bra-nap1    DFKLGETVMT G.PCKS.PIE FTLQGNVKT. .DGGST..QG KDRWV....V
ara-tha1    EYMVESLEFK G.PCKG.PVT LELNGNFKAP ATVKTT..KP HAGWI....D
zea-mayx    DFLVGQLNFT G.PCKG.DVT IQVDGNLLAT TDLSQY..KD HGNWI....E
ara-tha2    TFLAGPVIFA G.PCKS.KVT VNVIGTIIAT TSGYAT..PE ...WF....L
```

```
            201                                                    250
erw-caro    CG.VVDKNGK GCDAFITAVS TTNSGIYGPG TIDGQGGVKL ....QDKKVS
pse-sola    CGTATTSNDK SCNALIVARD TAGSGIVGAG AIDGRGGSLV TSGPNANRLT
asp-fla2    RFGGNKITVT QA........ ........AA VIDVQGSRWW DGKG..PN..
asp-nig1    RFGGKDLTVT MA........ .......DGA VIDGDGSRWW DSKG..TN..
asp-nig2    SMSGEHITVT GA........ .......SGH LINCDGARWW DGKG..T...
asp-tub2    SMSGKDITVT GA........ .......SGH LINCDGARWW DGKG..T...
 asp-ory    SVSGTNIKVQ QA........ .......SGA KIDGDGSRWW DGKG..GN..
asp_fla1    SVSGTNIKVQ QA........ .......SGA KIDGDGSRWW DGKG..GN..
 asp-par    SVSGTNIKVQ QA........ .......SGA KIDGDGSRWW DGEG..GN..
 col-lin    SFSGSNVVID GA........ .......SGH SIDCQGSRWW DSKG..GN..
 crp-par    SVSGTDIEVT GA........ .......SGH VIDGNGAAWW DGEG..SN..
coc-car1    SVSGTNIKVV GA........ .......SGH TIDAAGQKWW DGKG..SN..
scl-scle    SVSGTDITVT GT........ .......SGS VLDGNGAKYW DGKG..TN..
asp-nidu    RVSGTDITVE GE........ .......SDA VLNGDGSRWW DGEG..GN..
 fus-mon    VISGSNITIT GA........ .......SGH VIDGNGQAYW DGKGSNSN..
chon-pur    TIDGSNIKFK GA........ ........GH KFDGNGAKYW DGQG..TN..
asp-tub1    FFQLGGEDVN MY........ .......GGG TINGNGQVWY DLYAEDDL..
coc-car2    FFQLGGKDIN VY........ .......GGG TLDGNGQAWY DLYAKDIY..
mal-dome    FDNVQNLLVV G......... ........PG TINGNGNIWW ..KNSCKI..
pru-per2    FDNVQSLLVV G......... ........PG TINGNGNRWW ..ENSCKR..
bra-nap2    LEDVNNLSID GG........ .......SAG IVDGNGKIWW ..QNSCKI..
lyc-esc1    FDSVQNLVVG GG........ .........G TINGNGQVWW ..PSSCKI..
per-amer    FEDISNLTLE GG........ .........G TSNGNGETWW ..DSSCKR..
lyc-esc2    FEKVNGLSIY G......... .........G TFDGQAAALW ACKNSNNK..
pru-per1    FHHVNGVTIS G......... .........G ILDGQGTALW ACKACHGE..
 gos-bar    FYSVENFKMF GG........ .........G IFDGQGSIAY EKNTCE....
 gos-hir    FYSVENFKMF GG........ .........G IFDGQGSIAY EKNTCE....
nic-taba    VNKLDQFTMS GG........ .........G ILDGQAAAAW ECKQSK....
 oen-org    LNKIDLLTIF GG........ .........G VFDGQGKSAW VQNDCH....
 med-sat    FLYMDHLTLS GK........ .........G VFDGQGATVY KKAAPASA..
bra-nap1    FEKINGFKLN GG........ .........G TFDGEGNAAW KANNCHKT..
ara-tha1    FENIADFTLN GN........ ........KA IFDGQGSLAW KANDCAKT..
zea-mayx    ILRVDNL.VI TG........ ........KG NLDGQGPAVW SKNSCTKK..
ara-tha2    FERVDNLVL. TG........ ........TG TFHGKGEAVW KADGCGKK..
```

```
           251                                                     300
erw-caro   WWELAADAKV KKLKQNTPRL IQINKSKNFT LYNVSLINSP NFHVVFSDGD
pse-sola   WWDIAYLNKT KGLNQQNPRL IQTYNGSAFT LYGVTVQNSP NFHIVTTGTS
asp-fla2   .......GGK TKPKFIQYPQ LE....S.PT ITGLHVKNSP VQVFSVQGN.
asp-nig1   .......GGK TKPKFMYIHD VE....D.ST FKGINIKNTP VQAISVQAT.
asp-nig2   .......SGK KKPKFFYAHG LD....S.SS ITGLNIKNTP LMAFSVQAN.
asp-tub2   .......SGK KKPKFFYAHG LD....S.SS ITGLNIKNTP LMAFSVQAD.
 asp-ory   .......GGK TKPKFFYAHK LD....S.SS ITGLQIYNTP VQGFSIQSD.
asp_fla1   .......GGK TKPKFCYVHK LD....S.SS ITGLQIYNTP VQGFSIQSD.
 asp-par   .......GGK TKPKFFYAHK LD....S.SS ITGLQIYNTP VQGFSIQSD.
 col-lin   .......GGK TKPKFFYAHS LK....D.ST IRGLHTLNTP VQAFSINGAA
 crp-par   .......GGK TKPKMFYAHS LK....Q.ST IHNLKVKNTP VQFMSINSAT
coc-car1   .......GGK TKPKFFYAHS LT....T.SS ISGLNIKNTP VQAFSINGVT
scl-scle   .......GGK TKPKFFYAHS LK....GKSS INNVKILNSP VQGFSINSAS
asp-nidu   .......GGK TKPKFFYAHD LT....S.ST IKSIYIENSP VQVFSIDGST
 fus-mon   .......SNQ KPDHFIVVQK TT....GNSK ITNLNIQNWP VHCFDITGSS
chon-pur   .......GGV TKPHPFL..K FK....GSGQ YSSFTVLNSP AQAISIGNSD
asp-tub1   .......ILR .......PIL MGIIGLNGGT IGPLKLRYSP QYYHFVANSS
coc-car2   .......ILR .......PIL FGLIGAKNAK ISDLKFRYSP QWYTLVANSS
mal-dome   .......KPQ PPCGTYAPTA VTFNRCNNLV VKNLNIQDAQ QIHVIFQNCI
pru-per2   .......KPQ PPCNEQAPTA VTFNKCNNLV VKNLKIQDAQ QMHVRFQNCK
bra-nap2   .......DKS KPC.TKAPTA LTLYNLNNLN VKNLRVRNAQ QIQISIEKCN
lyc-esc1   .......NKS LPCRD.APTA LTFWNCKNLK VNNLKSKNAQ QIHIKFESCT
per-amer   .......KKS LPCKS.APTA LTFRSCKNLI VSDLSIKDSQ KMHLSFDKCQ
lyc-esc2   .......... ..NCPDGTTA LTFYNSNNII MSGVKVQNSQ KFQILVDGCH
pru-per1   .......... ..SCPSGATT LGFSDSNNIV VSGLASLNSQ MFHIVINDFQ
 gos-bar   ........NR E.FRSKLPVN IRFDFVTNAL IQDITSKDSK LFHINVFACK
 gos-hir   ........NR E.FRSKLPVN IRFDFLTNAL IQDITSKDSK LFHINVFACK
nic-taba   ........K. ...CNKLPNN LSFNSLTNST IKDITTLDSK SFHVNVNQCK
 oen-org   ........KN GPICKTLSMN LRLYAVTNSI LRDVTTLDSK NFHVNVIGCK
 med-sat   .......WSG KNSNSKVFMN FGFNFVNNSI VRGVTSKDSK NFHVMVFGCK
bra-nap1   .......FE. ...CKKLPIS VRFDFVDNAE IKDVTSLDAK NFHFNVISGK
ara-tha1   .......GK. ...CNSLPIN IRFTGLTNSK INSITSTNSK LFHMNILNCK
zea-mayx   .......YD. ...CKILPNS LVMDFVNNGE VSGVTLLNSK FFHMNMYRCK
ara-tha2   .......VQ. ...CNLPPTS LKFRNMKNVE INGISSVNAK AFHMFLVKTE
```

```
          301                                                350
erw-caro  GFTAWKTTIK TPS....... .......... .......... .TARNTDGID
pse-sola  GVTAWGIKIV TPSLAYAVAG YKCPSGSTPD KVTPATCFTP ETVKNTDGFD
asp-fla2  DVHLTDITID NSD....... .........G D.......NN G.GHNTDAFD
asp-nig1  NVHLNDFTID NSD....... .........G D.......DN G.GHNTDGFD
asp-nig2  DITFTDVTIN NAD....... .........G D.......TQ G.GHNTDAFD
asp-tub2  DITLTDITIN NAD....... .........G D.......TL G.GHNTDAFD
 asp-ory  NLNITDVTID NSA....... .........G ........TA E.GHNTDAFD
asp_fla1  NLNITDVTID NSA....... .........G ........TA E.GHNTDAFD
 asp-par  NLNITDVTID NSA....... .........G ........TA E.GHNTDAFD
 col-lin  NLGVYDVSVD NSA....... .........G D.......SA G.GHNTDAFD
 crp-par  DLNVIDVTMD NSA....... .........G A.......SK ..GHNTDAFD
coc-car1  GLTLDRITID NSA....... .........G D.......SA G.AHNTDAFD
scl-scle  GLTLSGITID NSA....... .........G N.......SL ..GHNTDAFD
asp-nidu  DLTMTDITVD NTD....... .........G D.......TD DLAANTDGFD
 fus-mon  QLTISGLILD NRA....... .........G DKPNAKSGSL PAAHNTDGFD
chon-pur  GLTFDTVTVD NSA....... .........G D.......SG SLGHNTDGFD
asp-tub1  NVLFDGIDIS GYS....... .........K ........SD NEAKNTDGWD
coc-car2  QVVFSNIDIF GDS....... .........K ........SK NPAKNTDGWD
mal-dome  NVQASCLTVT APE....... .........D ........S. ...PNTDGIH
pru-per2  NVEASHLTVT APE....... .........D ........S. ...PNTDGIH
bra-nap2  SVDVKNVKIT APG....... .........D ........S. ...PNTDGIH
lyc-esc1  NVVASNLMIN ASA....... .........K ........S. ...PNTDGVH
per-amer  DVIASNLMVT APE....... .........H ........S. ...PNTDGIH
lyc-esc2  NVKLQGVKVS APG....... .........N ........S. ...PNTDGIH
pru-per1  NVQMQGVRVS RSG....... .........N ........S. ...PNTDGIH
 gos-bar  NITLERLKIE APD....... .........E ........S. ...PNTDGIH
 gos-hir  NITLERLKIE APD....... .........E ........S. ...PNTDGIH
nic-taba  NLTFIRFNVS APA....... .........N ........S. ...PNTDGIH
 oen-org  NLTFERFKIS AAE....... .........T ........S. ...INTDGIH
 med-sat  NITFDGFTIT APG....... .........D ........S. ...PNTDGIH
bra-nap1  NMTFDNIKII APA....... .........E ........S. ...PNTDGIH
ara-tha1  NITLSDIGID APP....... .........E ........S. ...LNTDGIH
zea-mayx  DMLIKDVTVT APG....... .........D ........S. ...PNTDGIH
ara-tha2  NVNIQNIKLT APA....... .........E ........S. ...PNTDGIH
```

```
           351                                                  400
erw-caro   PMSSKNITIA YSNIATGDDN VAIKAYKGRA ETRNISILHN DFGTGHGMSI
pse-sola   PGQSTNVVLA YSYINTGDDH VAVKASSGP. .TRNLLFAHN HFYYGHGLSI
asp-fla2   VSESNGVYIT GANVKNQDDC LAINS..... GE.NIEFTGA TCSGGHGISI
asp-nig1   ISESTGVYIS GATVKNQDDC IAINS..... GE.SISFTGG TCSGGHGLSI
asp-nig2   VGNSVGVNII KPWVHNQDDC LAVNS..... GE.NIWFTGG TCIGGHGLSI
asp-tub2   VGNSVGVNII KPWVHNQDDC LAINS..... GE.NIWFTSG TCIGGHGLSI
 asp-ory   VGSSTYINID GATVYNQDDC LAINS..... GS.HITFTNG YCDGGHGLSI
asp_fla1   VGSSTYINID GATVYNQDDC LAINS..... GS.HITFTNG YCDGGHGLSI
 asp-par   IGSSTYINID GATVYNQDDC LAINS..... GS.HITFTNG YCDGGHGLSI
 col-lin   VGSSTGVYIS GADVKNQDDC LAVNS..... GT.NITFTGG TCSGGHGLSI
 crp-par   VGSSENIYIS GAVINNQDDC LAINS..... GT.NITFTSG SCTGGHGLSI
coc-car1   IGSSSGITIS NANIKNQDDC VAINS..... GS.DIHVTNC QCSGGHGVSI
scl-scle   VGSSTDITIS GANVQNQDDC LAINS..... GT.GITFTGG TCSGGHGLSI
asp-nidu   IGESTYITIT GAEIYNQDDC VAINS..... GE.NIYFSAS VCSGGHGLSI
 fus-mon   ISSSDHVTLD NNHVYNQDDC VAVTS..... GT.NIVVSNM YCSGGHGLSI
chon-pur   V.SADNVTIQ NSVVKNQDDC IAIND..... RS.NIVFQNN QCSGGHGISV
asp-tub1   TYRSNNIVIQ NSVINNGDDC VSFKP..... NSTNILVQNL HCNGSHGISV
coc-car2   TYRSDNIIIQ NSNINNGDDC VSFKP..... NSTNILVQNL VCNGSHGISV
mal-dome   VTNTQNITIS SSVIGTGDDC ISIVS..... GSQRVQATDI TCGPGHGISI
pru-per2   ITNTKNITIS SSVIGTGDDC ISIVS..... GSQRVQATDI TCGPGHGISI
bra-nap2   IVATKNIRIS NSDIGTGDDC ISIED..... GSQNVQINDL TCGPGHGISI
lyc-esc1   VSNTQYIQIS DTIIGTGDDC ISIVS..... GSQNVQATNI TCGPGHGISI
per-amer   ITGTQRIHVM NSVIGTGDDC ISIES..... GSKMVIATNI TCGPGHGISI
lyc-esc2   VKSSSGVSIM KSQIGTGDDC ISIGP..... GTSNLWIEGI ACGPGHGISI
pru-per1   VQMSSGVTIL NSKIATGDDC VSIGP..... GTSNLWIEGV ACGPGHGISI
 gos-bar   MGKSEGVNII ASDIKTGDDC ISIGD..... GTKNMVIKEI TCGPGHGISI
 gos-hir   MGKSEGVNII ASDIKTGDDC ISIGD..... GTKNMVIKEI TCGPGHGISI
nic-taba   VSRSSSVNIT DSNFSTGDDC ISVGD..... ETEQLYITRV TCGPGHGISV
 oen-org   IGRSDGVNII NTEIKTGDDC ISLGD..... GSKNINITNI TCGPGHGISV
 med-sat   MGKSTDVKIL NTNIGTGDDC VSIGD..... GSKQITVQGV NCGPGHGLSV
bra-nap1   LGRCEGVKIL NTKIATGDDC ISVGD..... GMKNLLIEKV VCGPGHGISV
ara-tha1   IGRSNGVNLI GAKIKTGDDC VSIGD..... GTENLIVENV ECGPGHGISI
zea-mayx   MGDSSGITIT NTVIGVGDDC ISIGP..... GTSKVNITGV TCGPGHGISI
ara-tha2   LSNADNVSIL DSTIATGDDC VSVGR..... GSNNVTVERV ICGPGHGLSV
```

```
           401                                                  450
erw-caro   GSET.MGVYN VTVDDLKM.N G....TTNGL RIKS...... .......DKS
pse-sola   GSETNTGVSN MLVTDLTM.D GNDSSAGNGL RIKS...... .......DAS
asp-fla2   GSIGNRDTN. .TVKNVKVAD STVVDSDNGI RIKT...... ......ISGA
asp-nig1   GSVGGRDDN. .TVKNVTISD STVSNSANGV RIKT...... ......IYKE
asp-nig2   GSVGDRSNN. .VVKNVTIEH STVSNSENAV RIKT...... ......ISGA
asp-tub2   GSVGGRSNN. .VVKNVTIEH STVSNSENAV RIKT...... ......VSGA
 asp-ory   GSVGGRSDN. .TVEDVTISN SKVVNSQNGV RIKT...... ......VYDA
asp_fla1   GSVGGRSDN. .TVEDVTISN SKVVNSQNGV RIKT...... ......VYDA
 asp-par   GSVGGRSDN. .TVEDVTISN SKVVNSQNGV RIKT...... ......VYGA
 col-lin   GSVGGRKDN. .VVKSVSITN SKIINSDNGV RIKT...... ......VAGA
 crp-par   GSVGGRSDN. .TVKTVSITN SKIINSQNGV RIKT...... ......VYDA
coc-car1   GSVGGRKDN. .TVKGVVVSG TTIANSDNGV RIKT...... ......ISGA
scl-scle   GSVGGRSDN. .VVSDVIIES STVKNSANGV RIKT...... ......VSGA
asp-nidu   GSVGGRDDN. .TVKNVTFYD VNVLKSQQAI RIKT...... ......IYGD
 fus-mon   GSVGGKSDN. .VVDGVQFLS SQVVNSQNGC RIKS...... ......NSGA
chon-pur   GSIA..SGK. .HVSGVAIKG NTVTNSMYGM RVKV...... ......KAAA
asp-tub1   GSLGQYKDEV DIVENVYVYN ISMFNASDMA RIKVWPGTPS ALSADLQGGG
coc-car2   GSLGQYPGEV DIVENILVRN ISMSNASDGA RIKVWPGASS ALSGDLQGGG
mal-dome   GSLGEDGSE. DHVSGVFVNG AKLSGTSNGL RIKTWK.... .......GGS
pru-per2   GSLGEDNAN. DHVSGVFVNG AKISGTSNGV RIKTWQ.... .......GGS
bra-nap2   GSLGDDNSK. AYVSGINVDG ATLSETDNGV RIKTYQ.... .......GGS
lyc-esc1   GSLGSGNSE. AYVSNVTVNE AKIIGAENGV RIKTWQ.... .......GGS
per-amer   GSLGDRNSE. AHVSGVLVDG GNLFDTTNGL RIKTWQ.... .......GGS
lyc-esc2   GSLGWKQQE. LGVQNVTVKT VTFSGTTNGV RVKTWA.... .......RPS
pru-per1   GSLGKEQEE. AGVQNVTVKT VTFSGTQNGL RIKSWG.... .......RPS
 gos-bar   GSLGKFQNE. EPVEGIKISN CTITNTSNGA RIKTWP.... .......GEH
 gos-hir   GSLGKFQNE. EPVEGIKISN CTITNTSNGA RIKTWP.... .......GEH
nic-taba   GSLGGNPDE. KPVVGVFVRN CTFTNTDNGV RIKTWP.... .......ASH
 oen-org   GSLGRYKNE. ESVVGIYVKN CTITGSQNGV RIKTWP.... .......KSE
 med-sat   GSLGKFTTE. ENVEGITVKN CTLTATDNGV RIKTWP.... .......DAP
bra-nap1   GSLGRYGWE. QDVTDITVKN CTLEGTSNGL RIKTWP.... .......SAA
ara-tha1   GSLGRYPNE. QPVKGVTVRK CLIKNTDNGV RIKTWP.... .......GSP
zea-mayx   GSLGRYKDE. KDVTDINVKD CTLKKTMFGV RIKAYE.... .......DAA
ara-tha2   GSLGKYKNE. EDVSGIHVNN CTMIETDNGL RIKTWG.... .......GSD
```

```
           451                                                        500
erw-caro   AAGVVNGVRY SNVVMKNVAK .PIVIDTVYE KKEGSNVPDW SD...ITFK.
pse-sola   RGGKVTNIVY DGICMRNVKE .PLVFDPFYS SVKGSLYPNF TN...IVVKN
asp-fla2   T.GSVSGVTY ENITLKNIKK NGIVIELDY. .KNGGPTGKP TTGVPITDLT
asp-nig1   T.GDVSEITY SNIQLSGITD YGIVIEQDY. .ENGSPTGTP STGIPITDVT
asp-nig2   T.GSVSEITY SNIVMSGISD YGVVIQQDY. .EDGKPTGKP TNGVTIQDVK
asp-tub2   T.GSVSEITY SNIVMSGISD YGVVIQQDY. .EDGKPTGKP TNGVTITDVK
 asp-ory   T.GTVSNVKF EDITLSGITK YGLIVEQDY. .ENGSPTGTP TNGIKVSDIT
asp_fla1   T.GTVSNVKF EDITLSGITK YGLIVEQDY. .ENGSPTGTP TNGIKVSDIT
 asp-par   T.GTVSNVKF EDITLSGITK YGLVVEQDY. .ENGSPTGTP TNGITVSGIT
 col-lin   T.GPVSDITY SGITLSNIAK YGIVIEQDY. .ENGSPTGKP TSGVPISGLT
 crp-par   T.GSVSDVTY SGITLSGITN YGIVIEQDY. .ENGSPTGTP TTGVPITGLT
coc-car1   T.GSVSDITY ENITLKNIAK YGIVIEQDY. .LNGGPTGKP TTGVPITGVT
scl-scle   T.GSVSGVTY KDITLSGITS YGVVIEQDY. .ENGSPTGKP TSGVPITGVT
asp-nidu   T.GSVSEVTY HEIAFSDATD YGIVIEQNY. .DDTSKT..P TTGVPITDFV
 fus-mon   T.GTINNVTY QNIALTNIST YGVDVQQDY. .LNGGPTGKP TNGVKISNIK
chon-pur   TSASVSAVTY SGNTISGIAK YGFLVSQSY. .PDDAST..P GTGAPSRGST
asp-tub1   GSGSVKNITY DTALIDNVDW .AIEITQCYG QKNTTLCNEY PSSLTISDVH
coc-car2   GSGAVRNVTY DGMIVKNVDY .AIEITQCYG QKNLTLCNQF PSNLTISDIT
mal-dome   GS..ATNIVF QNVQMNDVTN .PIIIDQNYC DHKTKDCKQQ KSAVQVKNVL
pru-per2   GS..ASNIVF QNVEMNDVTN .PIIIDQNYC DHKNKDCTRQ RSAVQVKNVL
bra-nap2   GT..AKNIKF QNIRMDNVKN .PIIIDQNYC D..KDKCEQQ ESAVQVNNVV
lyc-esc1   GQ..ASNIKF LNVEMQDVKY .PIIIDQNYC D.RVEPCIQQ FSAVQVKNVV
per-amer   GS..AKNIKF QNIVMHNVTN .PIIIDQYYC D.SKDPCPEQ ESAVKVSNVA
lyc-esc2   NGF.VRNVLF QHIVMVNVKN .PIIIDQNYC P.NHQSCPHK GSGIKISDVT
pru-per1   TGF.ARNILF QHATMVNVEN .PIVIDQHYC P.DNKGCPGQ VSGVQISDVT
 gos-bar   GG.AVSEIHF EDITMNNVSS .PILIDQQYC P.WNKCKKNE ESKVKLSNIS
 gos-hir   GG.AVSEIHF EDITMNNVSS .PILIDQQYC P.WNKCKKNE ESKVKLSNIS
nic-taba   PG.VVNDVHF EDIIVQNVSN .PVVIDQVYC P.FNKCNKDL PSQVKISKVS
 oen-org   PG.EASEMHF QDITMNSVGT .PILIDQGYC P.YNQCTAEV PSSVKLSKIS
 med-sat   GTITVSDIHF EDITMTNVKN .PVIIDQEYY P.WNQCSKKN PSKIKLSKIS
bra-nap1   CTTTAAGIHF EDIILNKVSN .PILIDQEYC P.WNQCNKNK PSTIKLVDIT
ara-tha1   PGIA.SNILF EDITMDNVSL .PVLIDQEYC P.YGHCKAGV PSQVKLSDVT
zea-mayx   SVLTVSKIHY ENIKMEDSAN .PIFIDMKYC P.NKLCTANG ASKVTVKDVT
ara-tha2   PSKAV.DIKF ENIIMQSVKN .PIIIDQNY. .....GSRGG DSQVAISDIL
```

```
          501                                                550
erw-caro  ..DVTSE... ..TKGVVVLN GENAKKPIEV TMKNVKLTSD .STWQIKNVN
pse-sola  FHDLGSAKSI KRTMTFLGYK ANKQKNPLTI TLDNVVFDGT LPAFEGSHYG
asp-fla2  VNGVTGSVAS KA.TPVYIL. .CGKGSCSDW TWKGVSISGG KKSDKCQNIP
asp-nig1  VDGVTGTLED DA.TQVYIL. .CGDGSCSDW TWSGVDLSGG KTSDKCENVP
asp-nig2  LESVTGSVDS GA.TEIYLL. .CGSGSCSDW TWDDVKVTGG KKSTACKNFP
asp-tub2  LESVTGTVDS KA.TDIYLL. .CGSGSCSDW TWDDVKVTGG KKSTACKNYP
asp-ory   FDKVTGTVES DA.TDIYIL. .CGSGSCTGL TWSGVSITGG KTSSKCENVS
asp_fla1  FDKVTGTVES DA.TDIYIL. .CGSGSCTDW TWSGVSITGG KTSSKCENVP
asp-par   FEKVTGTVES DA.TDIYIL. .CGSGSCTDW TWSGVSITGG KTSSKCENVP
col-lin   LSKISGSVSS SA.TPVYIL. .C..ASCTNW KWSGVSVTGG KKSSKCTGIP
crp-par   VSKVTGSVAS SA.TDVYIL. .CGKGSCSGW KWSGNSVTGG KKSSSCKNIP
coc-car1  LKNVAGSVTG SG.TEIYVL. .CGKGSCSGW NWSGVSITGG KKSSSCLNVP
scl-scle  LSNVHGTVSS SA.TNVYVL. .CAK..CSGW TW.DVNVTGG KTSTKCAGLP
asp-nidu  LENIVGTCED DDCTEVYIA. .CGDGSCSDW TWTGVSVTGG SVSDDCLNVP
fus-mon   FIKVTGTVAS SAQ.DWFIL. .CGDGSCSGF TFSGNAITGG GKTSSCNYPT
chon-pur  SVVILPISIN SGAKRVTVD. .C.RSCTGTW NWSKLTATGG SAGTISSDKA
asp-tub1  IKNFRGTTSG SEDPYVGTIV CSSPDTCSDI YTSNINVTSP DGTND.....
coc-car2  IKNFKGTTSK KYDPRVGYVV CSSPKVCSDI SIENIDVKSP SGTNL.....
mal-dome  YQNIRGTSAS GD...AITLN CSQSVPCQGI VLQSVQL... ..QNGR....
pru-per2  YQNIRGTSAS TD...AITFN CSQSVPCQGI VLQNIQL... ..QNAR....
bra-nap2  YRNIQGTSAT DV...AIMFN CSVKYPCQGI VLENVNI... ..KGGK....
lyc-esc1  YENIKGTSAT KV...AIKFD CSTNFPCEGI IMENINLVGE SGKPSE....
per-amer  YMNIRGTSAS EV...AVKFD CSKSSPCQGY IVGNINLVGN GGKETT....
lyc-esc2  YQDIHGTSAT EV...AVKLD CSKSNPCSGI TLEDVNLSYQ NQQTE.....
pru-per1  YEDIHGTSAT EV...AVKFD CSPKHPCREI KLEDVKLTYK NQAAE.....
gos-bar   FKNIRGTSAL PE...AIKFI CSGSSPCQNV ELADIDIQHN G.AEPAT...
gos-hir   FKNIRGTSAL PE...AIKFI CSGSSPCQNV ELADIDIKHN G.AEPAT...
nic-taba  FQNIKGTSRT QD...AVSLL RSKGVPCEGI EVGDIDITYS GKEGPAK...
oen-org   FKNIKGTSTT KE...AVKLV CSKSFPCNGV ELADIDLTYS GKGGPAT...
med-sat   FKNVKGTSGT AE...GVVLI CSSAVPCDGV ELNNVDLKFN G..APTT...
bra-nap1  FRNIRGTSEN KD...AVKLL CSKGHPCENV EIGDINIEYT GPDGPPT...
ara-tha1  IKGIKGTSAT KV...AVKLM CSKGVPCTNI ALSDINLVHN GKEGPAV...
zea-mayx  FKNITGTSST PE...VVSLL CTAKVPCTGV TMDDVNVEYS GTNNKTM...
ara-tha2  FKNIRGTTIT KD...VVQIM CSKSVPCQGV NVVDVNLDYV GKTGGEKKSS
```

```
          551                                                 600
erw-caro  VKK....... .......... .......... .......... ..........
pse-sola  GPASPNGVHF TFGGTGPVSF ADAIVTSSTT DVTVTGTPGT AAAVDCSKAF
asp-fla2  ..SGASC... .......... .......... .......... ..........
asp-nig1  ..SGASC... .......... .......... .......... ..........
asp-nig2  ..SVASC... .......... .......... .......... ..........
asp-tub2  ..SVASC... .......... .......... .......... ..........
 asp-ory  ..TGASC... .......... .......... .......... ..........
asp_fla1  ..TGASC... .......... .......... .......... ..........
 asp-par  ..TGASC... .......... .......... .......... ..........
 col-lin  ..SGSGAACM VSYLFVLGA. .......... .......... ..........
 crp-par  ..SGASC... .......... .......... .......... ..........
coc-car1  ..SGASCMVA YALTSML... .......... .......... ..........
scl-scle  ..SGVKC... .......... .......... .......... ..........
asp-nidu  ..SGISCDL. .......... .......... .......... ..........
 fus-mon  ..NTCPS... .......... .......... .......... ..........
chon-pur  KISGGSY... .......... .......... .......... ..........
asp-tub1  ......FVCD NVDESLLSVN CTATSD.... .......... ..........
coc-car2  ......FTCA NAEGIQSQVN CTVEGDKGGH S......... ..........
mal-dome  ......AECN NVQPAYKGVV SPRC...... .......... ..........
pru-per2  ......AKCN NVKPAYKGAV SPRCSWGLVN .......... ..........
bra-nap2  ......ASCK NVNVKDKGTV SPKCP..... .......... ..........
lyc-esc1  ......ATCK NVHFNNAEHV TPHCTSLEIS EDEALLYNY. ..........
per-amer  ......MSCS NIVQGLLREG LSTFLFMKRR VHECSY.... ..........
lyc-esc2  ......ASCV NARGRVSGLQ KPTNCLLKS. .......... ..........
pru-per1  ......SSCS HADGTTEGVV QPTSCL.... .......... ..........
 gos-bar  ......SQCL NVKPITIGKL NPIPCSGPVP KTPSATA... ..........
 gos-hir  ......SQCL NVKPITSGKL NPIPCSGPVP KTPSATA... ..........
nic-taba  ......SSCE NIKPSLKGKQ NPPVCTASAA SSS....... ..........
 oen-org  ......SVCE NIKPTIKGKQ IPAICSGSAA KAA....... ..........
 med-sat  ......AKCT NVKPLVTG.. TAPVCQAPGA PAASTTATPA ASKTATPAAG
bra-nap1  ......FECT NVTPKLVGAQ NPKACVGPVV KAPGKE.... ..........
ara-tha1  ......SACS NIKPILSGKL VPAACTEVAK PGP....... ..........
zea-mayx  ......AICT NAKGSTKGCL KELACF.... .......... ..........
ara-tha2  SGGLVGALCD NANVIFGGKL SFPMCPK... .......... ..........
```

```
          601        612
erw-caro  .......... ..
pse-sola  VPLKSVAPTS PI
asp-fla2  .......... ..
asp-nig1  .......... ..
asp-nig2  .......... ..
asp-tub2  .......... ..
 asp-ory  .......... ..
asp_fla1  .......... ..
 asp-par  .......... ..
 col-lin  .......... ..
 crp-par  .......... ..
coc-car1  .......... ..
scl-scle  .......... ..
asp-nidu  .......... ..
 fus-mon  .......... ..
chon-pur  .......... ..
asp-tub1  .......... ..
coc-car2  .......... ..
mal-dome  .......... ..
pru-per2  .......... ..
bra-nap2  .......... ..
lyc-esc1  .......... ..
per-amer  .......... ..
lyc-esc2  .......... ..
pru-per1  .......... ..
 gos-bar  .......... ..
 gos-hir  .......... ..
nic-taba  .......... ..
 oen-org  .......... ..
 med-sat  KSPAK..... ..
bra-nap1  .......... ..
ara-tha1  .......... ..
zea-mayx  .......... ..
ara-tha2  .......... ..
```

Appendix 2

Towards transformation of *Chondrosterum purpureum*

Overexpression and disruption of endoPG gene requires an efficient transformation system for *C. purpureum*. Cotransformation was proposed as the primary selection for the endoPG gene engineered strains. To demonstrate that *C. purpureum* is transformable and to provide a readily usable selective marker gene, plasmids conferring antibiotic resistance in fungi were used to transform *C. purpureum*. Vectors containing promoters from ascomycetes driving the expression of hygromycin phosphotransferase, such as pAN-7 and pPS-57, failed to confer hygromycin resistance in *C. purpureum*.

The vector pGfAgt containing the *Streptoalloteichus hindustanus ble* gene regulated by the *Schizophyllum commune GPD* promoter and the *Agaricus bisporus GPD* terminator (Schuren, et al, 1994) appeared to transform *C. purpureum* to phleomycin resistance on the first try, but Southern hybridization showed that no stable transformtion has been achieved.

1. Method chosen

The method used in the transformation of *C. purpureum* is the one commonly used in fungal transformation. That is, to protoplast *C. purpureum* and then to introduce foreign DNA into the protoplasts in the presence of polyethyl glycol and calcium ions (Ficham, 1989). It is necessary to stably introduce additional copies of endoPG gene into the genome of *C. purpureum* for endoPG overexpression. This can be achieved by

cotransformation of the fungus with an appropriate selectable marker and the gene of interest. If two different plasmids are included during the transformation procedure but selection is applied for a gene present only on one of them, the other will also be found integrated in a high proportion of the resulting transformants (Timberlake, et al, 1985; Wernars, et al, 1987).

2. Protoplasting

Protoplasting was done according to the protocol developed for *Aspergillus niger*. Protoplasts were obtained from fresh mycelium grown for 5-10 days, chopped in a Warring blender and grown for another 10-14 days. It was found that the use of 100mg of the lysing enzymes per 1g of the mycelium (wet weight) and about 4-hour incubation of the mycelium with lysing enzymes at 25-30°C favored optimal protoplast yield.

Serial dilutions of the protoplasts were plated on regeneration medium containing 0.5 M $MgSO_4$ as osmotic stabilizer as well as on regeneration medium without osmotic stabilizer. The regeneration frequency of the protoplasts of *C. purpureum* was found to be very low as compared to some of the other fungi. The regeneration frequency of the protoplasts of *C. purpureum* was between 0.001/100 and 0.01/100. No regeneration of protoplasts was observed on medium without osmotic stabilizer demonstrating that regeneration from hyphal fragments did not occur.

3. Minimum Inhibitory concentration of phleomycin to *C. purpureum*

Protoplasts were plated on regeneration medium plus 0.5 M $MgSO_4$ with phleomycin at various concentrations to determine the minimum inhibitory concentration of phleomycin. Originally, 2.5 ug/ml was found to be the killing concentration at or above which no regeneration of the protoplasts was observed. However, later on, in two repeated experiments, the killing concentration seemed to be much higher, above 25ug/ml and 50ug/ml respectively. This inconsistency could be due to batch variation in the phleomycin, breakdown of antibiotic or an unknown difference in the fungus itself. The minimum inhibitory concentration still remains to be determined in future experiments.

4. Restriction analysis of the plasmid pGfAgt

Three plasmids, pGfAgt, pXAgft and pGft conferring phleomycin resistance, were obtained from Dr. Frank H.J. Schuren, U. of Groningen, The Netherlands. The plasmid pGfAgt was used in my transformation experiments. Transformation is usually by stable integration of the foreign DNA at several sites in the genome. It is necessary to know the restriction sites of the vector to predict hybridization patterns of putative transformants by Southern hybridization analysis (Ward, 1991). Six restriction enzymes were used singly and doubly to digest pGfAgt. It seemed that except for NcoI, which did not cut the plasmid, the other five - KpnI, BamHI, HindIII, EcoRI, and PstI, all had a sole site on the plasmid. In addition, SalI was also found to have a sole site on pGfAgt.

5. Transformation experiments

Transformation was done essentially according to either the protocol for the ascomycete fungus *Aspergillus niger* or the basidiomycete fungus *schizophylum commume* (Schuren, et al, 1988). *Chondrostereum purpureum* appeared to be transformed to phleomycin resistance using pGfAgt in one experiment. In the experiment, the minimum inhibitory concentration of phleomycin to untransformed *C. purpureum* was determined to be 2.5 ug/ml. The putative transformants were able to regenerate on plates containing 2.5 ug/ml and 25 ug/ml phleomycin while the untransformed was completely inhibited by 2.5 ug/ml phleomycin. The putative transformants were isolated and transferred twice onto plates containing phleomycin of the same concentration as the original plate. Thirteen of the putative transformants were analysed immediately after. DNAs were extracted from the mycelium inoculated in malt extract broth without phleomycin, digested with SalI and NcoI individually, size fractionated by gel electrophoresis and blotted onto nylon membranes. ^{32}P labeled pGfAgt (uncut) was hybridized to these blots. The Southern hybridization appeared to demonstrate integration events in all six transformants on one blot, but the signals were not clear and the background was somewhat high. No hybridization signals were detected on the other two blots, one containing another seven transformants and one control containing restricted genomic DNA of the wild type *C. purpureum*. The probes were stripped off from these three blots and the hybridization was repeated, but no signal was detected the second time. A year later, for a thorough conclusion, all the remaining seventeen putative transformants from this transformation kept on phleomycin plates at 4°C were transferred onto four fresh phleomycin plates with phleomycin concentration of 25 ug/ml, 50 ug/ml, 125 ug/ml and 250 ug/ml respectively.

Observation of their growth was recorded and appeared to be interesting. Different putative transformants seemed to grow differently on these plates: some maintained vigorous growth on even the highest concentration of phleomycin while others did not grow well on the lowest concentration of phleomycin. The wild type did not grow at all even on the lowest phleomycin concentration. DNAs were extracted from all of them for a Southern analysis. No specific hybridization signal was found. Hence, either the transformation was not really achieved or it was a transient (unstable) transformation event.

The transformation experiment was repeated six times more. Each time, fresh protoplasts were made. Four out of the six experiments were contaminated with bacteria or *Aspergillus.* For the two repeated transformations which were free of contamination, the minimum inhibitory concentration of phleomycin to *C. purpureum* seemed much higher than previously determined. Both times, regeneration of protoplasts, both from the transformed and the untransformed, were observed on all the plates containing phleomycin form 2.5 ug/ml to 50 ug/ml, making it impossible to determine if transformation was achieved or not in these experiments.

VITA

Surname: Tang　　　　　　Given name: Yijian

Place of birth: Xiangtan, Hunan, P.R.China

Educational institutions attended:

University of Victoria	1993-1997
Hunan Medical University	1984-1990

Degrees Awarded:

B.M. Hunan Medical University

Honors and Awards:

Graduate Teaching Fellowship, University of Victoria, 1993-1996

University Scholarship, Hunan Medical University, 1984-1987

Publications:

Ramsfield, T. D., Becker, E. M., Rathlef, S. M., Tang, Y., Vrain, T. C., Shamoun, S. F., and Hintz, W. E. 1996 Geographic variation of *Chondrostereum purpureum* detected by polymorphisms in the ribosomal DNA. Canadian Journal of Botany, 74: 1919-1929.

Partial Copy Right Licence

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission of extensive copying of this thesis for scholarly purpose may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

Cloning and molecular characterization of an polygalacturonase gene of the white-rot fungus *Chondrostereum purpureum*

Author

Yijian Tang

March, 1997